



TRABAJO FIN DE GRADO

GRADO EN LINGÜÍSTICA Y LENGUAS APLICADAS

**«Análisis estadístico – descriptivo de rasgos lingüísticos para
el reconocimiento del emisor»**

Autora: María Rey Millán

Tutor: D. Mario Crespo Miguel

Curso Académico 2014-15



Facultad de Filosofía y Letras

Septiembre 2015

Índice

1.	Resumen	3
2.	Introducción	5
	2.1. ¿Qué es la Lingüística Forense?	6
	2.2. Fonética Forense y Acústica Forense	8
	2.3. Características de un peritaje	9
3.	Solicitud	13
4.	Objetivos	13
5.	Identificación del perito y manifestaciones	14
6.	Muestras	15
	6.1. Calidad y validez del documento	15
	6.2. Muestras	16
7.	Estudio	18
	7.1. Fundamentos teóricos y prácticos	18
	7.1.1. Teóricos	22
	7.1.2. Herramientas	28
	7.1.3. Análisis	32
	7.2. Cotejo	33
	7.2.1. Análisis articulatorio	33
	7.2.2. Análisis acústico	43
	7.2.3. Análisis lingüístico	57
8.	Dictamen	62
9.	Bibliografía	68
10.	Anexos	71

1. Resumen

En esta investigación se ha llevado a cabo una combinación de trabajo académico y práctico. Hemos intentado seguir la estructura de los peritajes lingüísticos, que se consideraría la parte práctica antes mencionada, y le hemos añadido la teoría y la metodología necesaria para conocer todos los detalles del análisis realizado, que completarían la parte académica. El objetivo del trabajo es la identificación de locutores a partir de una serie de muestras que, en este caso, han sido tomadas por los autores del proyecto a alumnos de la Facultad de Filosofía y Letras de la Universidad de Cádiz, concretamente del grado de Lingüística y Lenguas Aplicadas. Aleatoriamente, hemos tomado una de ellas como la muestra dubitada y las otras como indubitadas y se les ha realizado una serie de análisis (perceptivo, articulatorio, acústico y lingüístico) para observar las características comunes y distintivas que existen entre ellas. Para el análisis articulatorio hemos tenido en cuenta los rasgos más característicos del andaluz, ya que nuestras muestras, al estar tomadas de la Universidad de Cádiz, cuentan con muchos de estos rasgos. Luego, se han analizado a partir de la realización de una transcripción. Además, al encontrarnos con un registro coloquial, nos hemos centrado en la utilización de vulgarismos. Hemos considerado este tipo de registro porque el hablante, al mantener una conversación con alguien cercano, se relaja a la hora de expresarse y tiende a cometer este tipo de “errores”. Por otro lado, en el análisis acústico, hemos decidido observar las medidas de los formantes 1 y 2 y del pitch de las vocales que denominamos fuertes, que son [a], [o] y [u]. A partir de las medidas del pitch, hemos realizado una comparación de la distancia euclídea y un análisis mediante Statdisk, con el que hemos podido comprobar, utilizando un test que nos proporcione el P-value, si los textos se parecen o, por el contrario, son diferentes. Por último, en el análisis lingüístico, hemos realizado el Chi – square Significance Test de las palabras y las letras de las transcripciones realizadas. Además se ha llevado a cabo un análisis con CopyCatch de la frecuencia del vocabulario de las muestras. Una vez realizados, hemos hecho una serie de comparaciones de todos los resultados con el fin de localizar, entre las muestras indubitadas, al autor de la muestra dubitada.

Palabras clave: Identificación de locutores, peritaje lingüístico, vocales, andaluz.

This research has been conducted a combination of academic and practical work. On the one hand we have structured as usually structured linguistic expertise, the practical part above would be considered, and secondly we have added the theory and methodology necessary to know all the details of the analysis, which would complete the academic part. The objective of this study is the identification of speakers from a number of samples, which in this case have been taken by the sponsors to students of the Faculty of Arts of the University of Cadiz, specifically the degree of Linguistics and Applied Languages. Randomly, we have taken one of them as the dubitable sample and the other as indubitable and have undergone a series of tests (perceptive, articulatory, acoustic and linguistic) to observe the common and distinctive features between them. On the one hand, to the articulatory analysis it has been observed the hallmarks of Andalusia, as our samples to be taken from the University of Cadiz, have many of these traits have been analyzed from the realization of a transcript. In addition, we have focused on the use of slang, as we found with some samples in a colloquial, making the speaker to talk and relax commits this kind of "errors". On the other hand, in the acoustic analysis, we decided to observe the measures formant 1 and 2 and the pitch of the voice we call strong, which are [a], [o] and [u]. From the measurements of the pitch, we made a comparison of the Euclidean distance. Finally, linguistic analysis, we performed the chi - square significance test of the words and letters of transcripts that we have made of the samples. It has also conducted an analysis CopyCatch frequency vocabulary samples and one with STATDISK, with which we have observed, through a test that will provide us the P-value, if the texts are similar or, conversely, they are different. Once done, we have made a series of comparisons of all the results in order to find, among the indubitable signs, the author of the dubitable sample.

2. Introducción

Esta investigación está planteada como una fusión entre un trabajo académico y un peritaje lingüístico. Nos ha parecido interesante este punto de vista para darle un toque un poco más profesional, ya que nos interesamos por las normas y el trabajo que realiza un perito en tareas lingüísticas. Respetaremos todo lo relevante que debe tener un trabajo académico y lo uniremos a un trabajo práctico poniéndonos en el papel de estos profesionales. El fin es simular un caso judicial. La parte académica, como cualquier trabajo de investigación, cuenta con una introducción, un resumen, unos objetivos que debemos alcanzar y la explicación de los fundamentos teóricos, donde explicaremos toda la información necesaria para este trabajo. También cuenta con un apartado donde daremos información sobre todas las herramientas que debemos utilizar para llevar a cabo los análisis pertinentes. Antes de realizar cualquier tipo de análisis, debemos explicar detalladamente de qué se trata y cómo los vamos a llevar a cabo. Además, debe tener una conclusión. Por otro lado tenemos el caso del peritaje, del que explicaremos cada una de sus partes a medida que avance esta investigación. El simulacro de peritaje que hemos realizado cuenta con una solicitud, donde se explica al perito el trabajo que debe realizar, una identificación del perito y un informe, donde debemos explicar las muestras con las que contamos y la validez de ellas. A esto le sigue un cotejo, que abarcará la realización de todos los análisis necesarios para obtener los resultados. Por último tenemos un dictamen, que en este caso, realiza la función de las conclusiones del trabajo académico. En este apartado se recopilan de los datos finales y se realiza una comparación de los mismos.

Centrándonos un poco más en el tema a tratar, el trabajo va a estar relacionado con la Lingüística Forense, y más concretamente en la identificación de hablantes. Para ello contaremos con una serie de muestras, que han sido tomadas manualmente por la autora de esta investigación. Se han recolectado un total de cinco muestras de alumnos de la Facultad de Filosofía y Letras de la Universidad de Cádiz, concretamente del grado de Lingüística y Lenguas Aplicadas.

Las grabaciones han sido realizadas con unos auriculares con micrófono integrado y mediante un programa informático llamado Praat. Este programa será explicado en los próximos epígrafes. Todas estas muestras que vamos a analizar, han sido tomadas en el mismo lugar y en las mismas condiciones. Estas pruebas las iremos detallando más detenidamente a medida que vamos avanzando en el estudio.

2.1. ¿Qué es la Lingüística Forense?

La Lingüística Forense se puede considerar una disciplina bastante novedosa en España, ya que su estudio es muy reciente en nuestro país. Pero dentro de los países anglosajones ya lleva más años en desarrollo. Es muy difícil proporcionar una definición exacta de esta disciplina. Lo que sí sabemos es que es multidimensional e interdisciplinar, donde entran expertos de todo tipo de materias que pueden ayudar a llevar a cabo las investigaciones oportunas realizadas por el lingüista. Algunos la colocan como la interfaz entre Lengua y Derecho, como es el caso de la Asociación Internacional de Lingüistas Forenses (IAFL). (Jiménez, M. et Al, 2012). Una de las definiciones más completas que hemos podido encontrar es la siguiente:

La Lingüística Forense es una técnica relacionada con una profunda y detallada evaluación de las características lingüísticas de un texto, incluyendo gramática, sintaxis, ortografía, vocabulario y fraseología, que se lleva a cabo mediante comparaciones de textos dubitados (autoría desconocida) e indubitados (de autoría conocida), en un intento de descubrir idiosincrasias peculiares de autoría a fin de determinar si los autores pueden o no ser el mismo. (Diccionario Legal de Black, 1990).

Debemos saber que un lingüista forense, según Shuy (2000), es el que aplica sus conocimientos y los pone al servicio de los demás para llevar a cabo un propósito forense. Deben aprenderse ante todo los principios y la práctica de las ciencias forenses, como pueden ser, por ejemplo, la defensa de análisis de evidencias lingüísticas en vistas orales o la determinación de la existencia de plagio en un texto escrito.

Podemos decir que, en nuestro país, los pioneros en avances e investigaciones son los profesionales de la Universidad Pompeu Fabra de Barcelona. Allí cuentan con un área destinada a la realización de peritajes lingüísticos, que es concretamente en lo que vamos a centrar este trabajo. Para ello cuentan con un laboratorio de Lingüística Forense (ForensicLab). (Jiménez, M. et Al, 2012).

La mayoría de los trabajos realizados por los analistas forenses pueden ser, según Watt (2010) de dos tipos. Por un lado los de reconocimiento de emisor, y por otro lado, los de perfil lingüístico. Este último se lleva a cabo cuando ningún sospechoso ha sido identificado, como por ejemplo cuando se producen grabaciones de llamadas telefónicas. Con él lo que pretendemos es reducir la población de posibles sospechosos, identificándolos por los rasgos lingüísticos asociados a ciertos grupos geográficos o sociales o a algunas características de pronunciación inusuales o patologías. Esta investigación está principalmente apoyada en la Tesis de Carlos Delgado sobre la Identificación de hablantes en el ámbito forense (2001). Además de este trabajo, nos hemos ayudado de otros trabajos y documentos más recientes que nos ayudan a completar y actualizar toda la información necesaria.

Para realizar una identificación de hablantes es necesario el análisis de una serie de grabaciones de habla, que son las que se utilizarán como muestras. Según los estudios de Carlos Delgado (2012), los primeros científicos forenses comenzaron a utilizar el análisis de estas grabaciones con el fin de establecer relaciones de correspondencia con posibles emisores. De esto hace ya más de medio siglo. Hasta prácticamente el año 2000, han existido diferentes aproximaciones metodológicas para llevar a cabo un análisis y un estudio de estas muestras. Para llevar a cabo un análisis de habla es un factor muy importante la experiencia que tenga el investigador y las alternativas de análisis. Además, debemos contar con unas muestras de similar duración y de buena calidad. Esto nos ayudará a realizar un estudio para obtener un nivel de certeza óptimo. (Delgado, C. 2012).

Existen varias herramientas específicas y especializadas para el fin al que pretendemos llegar. Como ejemplo de estas herramientas tenemos el programa BATVOX que, según Beatriz Gómez Sigüenza (2008), es un sistema automático para el reconocimiento de voz. A partir de este se han desarrollado otros programas para diferentes laboratorios forenses, como por ejemplo el Forensic Automatic Speaker Recognition Program (FASR). (Jiménez, M et Al.

2012). El programa BATVOX, al ser de pago, no se encuentra al alcance de todo el mundo para realizar este tipo de investigaciones.

Nosotros utilizaremos otro programa, llamado Praat, para realizar las mediciones, pero lo explicaremos en los próximos puntos, cuando hablemos de la metodología específica utilizada para llevar a cabo la investigación.

2.2. Fonética Forense y Acústica Forense

La Fonética Forense y la Acústica Forense son disciplinas pertenecientes a la Criminalística, que engloban la aplicación de técnicas desarrolladas por la ingeniería acústica para el esclarecimiento de delitos y la averiguación de la identidad de quienes los cometen. (Lucena Molina, J.J. 2005).

Estas disciplinas son las que más nos van a ayudar a la hora de realizar los análisis pertinentes. Sabemos que parten de las aportaciones de la Fonética y la Fonología sobre las diferentes lenguas y el estudio del lenguaje en general. La identificación de hablantes se lleva a cabo mediante dos subdisciplinas que podemos encontrar dentro de esta. Ellas se centran en el estudio de dos aspectos diferentes, que son, la fonética acústica y la articulatoria. Ambas han partido del estudio de mecanismos de habla. Estudian lo que hay en común entre todos los hablantes. (Lucena Molina, J.J. 2005). Según A. Quilis en su libro *Tratado de Fonología y Fonética Españolas* (1999), debemos distinguir, como hemos dicho anteriormente, dos subdisciplinas. La fonética articulatoria es la encargada de la articulación y la producción de sonidos. Las diferencias de articulación dependen de la distribución de los órganos que forman el aparato fonador, y que se encargan de esta producción. La composición del aparato fonador aparece indicada en el punto 7.1.1 de nuestro trabajo.

En definitiva, esta fonética estudia los cambios que sufre el aire que proviene de los pulmones cuando pasa por el aparato fonador, para convertirse en sonidos propios de la comunicación humana. (Quilis, A. 1999).

La fonética acústica por su parte es la encargada de las propiedades físicas de los sonidos articulados del habla humana durante la transmisión de un mensaje y la recepción que tiene el oyente de ella, es decir, la repercusión que tiene su percepción en el oído humano. Para que se pueda efectuar dicha transmisión es necesaria la existencia de un medio portador. El

sonido se propaga a través del aire a una velocidad de 340 metros por segundo aproximadamente. (Quilis, A. 1999).

Hay datos muy útiles y que debemos tener muy en cuenta para el análisis de las muestras, como son la determinación del origen geográfico de un hablante, la edad, el ámbito social del individuo, algunos rasgos de su personalidad y algunos rasgos patológicos que lo caractericen. (Lucena Molina, J.J. 2005).

Según recoge José Juan Lucena Molina (2005), Comandante y Jefe del departamento de Acústica e Imagen del Servicio de Criminalística de la Guardia Civil en un artículo para el Instituto Universitario de Investigación sobre Seguridad Interior, la asociación internacional en Europa que cuenta con los mejores expertos en lingüística dentro del campo del reconocimiento de locutores por el habla fue fundada en York, Inglaterra, en 1991 y es conocida como IAFPA (International Association for Forensic Phonetics and Acoustics).

2.3. Características de un peritaje

En esta investigación seguiremos los pasos que se deben dar para llevar a cabo un peritaje. Para ello debemos saber, en primer lugar cuál es el trabajo y el fin de un perito judicial y, en segundo lugar, cuál es la estructura con la que cuenta un informe sobre un peritaje lingüístico realizado. El perito judicial debe cumplir sus funciones bajo los siguientes principios de (Asociación Independiente de Peritos Judiciales):

- Razonabilidad. Deben establecer y exponer los fundamentos científicos que sustentan el informe.
- Congruencia. El informe debe guardar relación entre las premisas y las conclusiones.
- Proporcionalidad. Se deben establecer relaciones entre los hechos y sus consecuencias.
- Buena fe. No se deben utilizar pruebas ni artulugios que lleven a engaño a las partes o al juez en el proceso.
- No se puede alterar el supuesto bajo las cuales fue convocado, siendo su obligación principal, que es la de entregar el dictamen.

- Todos los materiales utilizados deben ser entregados para ayudar a la comprensión del informe por las partes y el magistrado.
- El perito judicial no es un testigo, sino que es un reconstructor de algún hecho o fenómeno.

No hay acuerdo en Lingüística forense sobre un índice de confianza estándar respecto a la certeza y a cómo deberían expresarse las conclusiones. Existen tres modos principales para expresar los resultados del análisis y comparación, que son: Escalas de certeza, probabilidades y verosimilitud. Esta última plantea una serie de problemas, que fueron expuestos por Coulthard, M en Coulthard, M. Y Johnson, A. (2007). Estos autores consideran que es necesario tener una buena base de datos sobre la que computar probabilidades. Ni los expertos forenses ni los juzgados desean dejar expresar y usar los resultados de manera categorial, es decir, con escalas. Esto es debido a que los expertos no saben cuantificar adecuadamente sus hallazgos. Las propuestas de Razón de verosimilitud (LR) no han sido refinadas adecuadamente. Existe un grado alto de complejidad inherente en variación entre muestras de habla natural. Desde el punto de vista de los destinatarios del trabajo, la interpretación de resultados como ratios de verosimilitud suele reportar más confusión que claridad. (Coulthard, M. Y Johnson, A, 2007)

Por su parte, Carlos Delgado, en 2001, establece que la determinación de los márgenes que delimitan la posible variación intra – e interpersonal han de ser establecidos de antemano a través de estudios experimentales o trabajos de campo de alta muestra.

Según el art. 346 LEC:

El Perito Judicial que el tribunal designe emitirá por escrito su dictamen, que hará llegar al tribunal en el plazo que se le haya señalado. De dicho dictamen se dará traslado a las partes por si consideran necesario que el perito concurra al juicio o a la vista a los efectos de que aporte las aclaraciones o explicaciones que sean oportunas. El tribunal podrá acordar, en todo caso, mediante providencia, que considera necesaria la presencia del perito en el juicio o vista para comprender y valorar mejor el dictamen realizado.

Siguiendo las pautas de la Asociación Independiente de Peritos judiciales, un peritaje lingüístico debe estar estructurado de la siguiente manera:

- I. Asunto
- II. Solicitud
- III. Finalidad
- IV. Identificación del perito y manifestaciones. Donde debe aparecer una justificación de la valía del perito y su experiencia. También se debe explicar la naturaleza del peritaje, su origen, su finalidad y sus fundamentos de actuación.
- V. Documentos. Se debe realizar una descripción de las muestras y los corpus con los que se cuenta. Son necesarias 30 o más observaciones, aunque se puede realizar un tamaño muestral estadístico. (Gigerenzer et al., 1982, cit En Rose, P. 2003).

Estas muestras deben seguir unos parámetros de análisis. Nolan (1983, cit. P. Rose, 2002) determina que los ideales son aquellos que muestran mucha variabilidad entre hablantes y poca variabilidad para un determinado individuo, que los rasgos no se puedan modificar voluntariamente, que no tengan gran frecuencia de aparición, que sean fácilmente extraíbles y medibles y que sean independientes de otros parámetros.
- VI. Estudio realizado. Debe contar con las explicaciones sobre los métodos que van a ser utilizados y su aceptación entre los científicos. Se debe especificar también cuál es el tipo de análisis que se va a llevar a cabo y las herramientas utilizadas, junto con su validez científica.

- VII. Dictamen. El peritaje trata de determinar la consistencia del texto indubitado oral o escrito respecto a otros textos precedentes del mismo autor, y respecto al perfil lingüístico esbozado. El resultado de la comparación debería ser una opinión fundada que ayudará al proceso legal.

La mayoría de los lingüistas forenses y especialistas del campo se han sentido tradicionalmente incapaces de expresar sus hallazgos en términos de probabilidades calculadas matemáticamente. Algunos expertos incluso llegan a expresar su opinión sin dar ninguna indicación sobre su “validez” y más allá de toda duda razonable. (Coulthart, 2010).

Existen varios tipos de peritajes lingüísticos, como pueden ser los realizados por plagio, determinación de autorías, perfiles lingüísticos, etc. (Jiménez, M. et Al. 2002). Nuestro trabajo se basa, como hemos dicho anteriormente, en la realización de un peritaje lingüístico, centrado en la identificación de locutores.

3. Solicitud

Esta investigación ha sido solicitada por la Facultad de Filosofía y letras de la Universidad de Cádiz (UCA) en su área de Lingüística y Lenguas Aplicadas y tutorizada por D. Mario Crespo Miguel, profesor de dicha área. Se trata de un trabajo de fin de grado, con el que se pretende obtener 6 créditos CTS necesarios para obtener el título del grado mencionado anteriormente. Es por ello por lo que el peritaje que se va a presentar cuenta con una parte académica, en la que se explica la teoría necesaria para poder realizarlo. Para el apartado práctico se han tomado un total de cinco muestras, con el fin de realizar los análisis pertinentes para lograr una identificación de locutores. Para ello se realizarán cuatro tipos de análisis, como son el perceptivo, el articulatorio, el acústico y el lingüístico, que explicaremos detalladamente a medida que avance el trabajo.

4. Objetivos

La finalidad de este proyecto es la de realizar un simulacro de un peritaje lingüístico con el objetivo de realizar una identificación de locutores, como ya hemos mencionado anteriormente. Tras la realización de esta investigación debemos haber aprendido a tratar grabaciones de voz con programas destinados al análisis y medición de las características necesarias para conocer los rasgos comunes y distintivos de las distintas muestras indubitadas y saber compararlas con los rasgos tomados de la muestra dubitada. El fin es deducir si alguno de los sospechosos puede ser el autor de la grabación principal.

5. Identificación del perito y manifestaciones

Doña María Rey Millán, graduada en Lingüística y Lenguas Aplicadas por la Universidad de Cádiz. Cuenta con conocimientos sobre Fonética y Fonología, Semántica y Sociolingüística, que son algunas de las ramas principales que se debe dominar para la realización de este proyecto. También cuenta con amplios conocimientos sobre Lingüística Forense, que será la base principal para la investigación. Atendiendo a lo contemplado en el art. 335.2 de Ley de Enjuiciamiento Civil, el perito actuante promete llevar a cabo el presente estudio con la mayor objetividad posible y con el único fin de averiguar la verdad y emitir el correspondiente dictamen, tomando en consideración tanto lo que pueda favorecer como lo que sea susceptible de causar perjuicio a cualquiera de las partes afectadas, indicando además que conoce el contenido de los artículos 458¹, 459² y 460³ del Código Penal.

La realización del estudio pericial se coloca dentro de un marco académico, con el fin de llevar a cabo un Trabajo Fin de Grado (TFG) solicitado por la Universidad de Cádiz para la consecución del título de Graduada en la carrera antes mencionada. El fin de dicho trabajo es el de realizar un análisis estadístico – descriptivo de rasgos lingüísticos para el reconocimiento del emisor.

¹ El testigo que faltare a la verdad en su testimonio en causa judicial, será castigado con penas de prisión de seis meses a dos años y multa de tres a seis meses. Si el falso testimonio se diera en contra del reo en causa criminal por delito, las penas serán de prisión de uno a tres años y multa de seis a doce meses. Si a consecuencia del testimonio hubiera recaído sentencia condenatoria, se impondrán las penas superiores en grado. Las mismas penas se impondrán si el falso testimonio tuviera lugar ante Tribunales Internacionales que, en virtud de Tratados debidamente ratificados conforme a la Constitución Española, ejerzan competencias derivadas de ella, o se realizara en España al declarar en virtud de comisión rogatoria remitida por un tribunal extranjero. (BOE-A-1995-25444).

² Las penas de los artículos precedentes se impondrán en su mitad superior a los peritos o intérpretes que faltaren a la verdad maliciosamente en su dictamen o traducción, los cuales serán, además, castigados con la pena de inhabilitación especial para profesión u oficio, empleo o cargo público, por tiempo de seis a doce años. (BOE-A-1995-25444).

³ Cuando el testigo, perito o intérprete, sin faltar sustancialmente a la verdad, la alterare con reticencias, inexactitudes o silenciando hechos o datos relevantes que le fueran conocidos, será castigado con la pena de multa de seis a doce meses y, en su caso, de suspensión de empleo o cargo público, profesión u oficio, de seis meses a tres años. (BOE-A-1995-25444).

6. Muestras

El trabajo se basa, como hemos dicho anteriormente, en el análisis de una serie de muestras tomadas personalmente por el investigador. Han sido grabadas a través del grabador del programa informático Praat (<http://www.fon.hum.uva.nl/praat/>), que explicaremos más adelante, ya que es el que utilizaremos posteriormente para llevar a cabo los análisis pertinentes con el fin de lograr el objetivo marcado en este trabajo. Hemos utilizado unos cascos con micrófono integrado para la recopilación de cinco muestras.

Debemos saber que, para realizar una identificación de hablantes, tenemos que tener una muestra dubitada y una serie de muestras indubitadas. Llamamos muestra dubitada a aquella que es desconocida, de la que queremos descubrir su autoría, mientras que las muestras indubitadas son las que tomamos conociendo a su autor. La identificación de locutores permite determinar de una manera bastante fiable si una persona ha emitido o no un determinado mensaje. Una vez que las muestras dubitadas llegan al laboratorio, se produce un estudio de calidad de las mismas. Más tarde se realiza una toma de muestras indubitadas al supuesto autor. Estas serán cotejadas mediante el uso de distintos tipos de análisis. Para ello nos ayudaremos de programas informáticos que nos ayudarán a tomar las mediciones pertinentes y se elaborarán los informes necesarios. (Jiménez, M et Al. 2012).

6.1. Calidad y validez del documento

Las muestras han sido tomadas de manera aleatoria entre los alumnos de cuarto curso del grado de Lingüística y Lenguas Aplicadas de la Facultad de Filosofía y Letras de la Universidad de Cádiz. A todos los individuos se les ha grabado mientras se les realizaba una breve entrevista de solo dos preguntas. De estas dos preguntas, la primera era una respuesta bastante simple y en la segunda se les daba opción a explicarse y extenderse todo lo que desearan, poniendo un límite de aproximadamente un minuto para que, a la hora de analizar las muestras, tuviéramos rasgos suficientes pero sin excedernos demasiado, ya que, al tratarse de un trabajo académico, tenemos ciertas limitaciones de extensión. Todas se han grabado en las mismas condiciones por los autores de esta investigación. Luego, mediante un sistema de elección aleatorio, hemos decidido cuál de las cinco muestras será la dubitada.

6.2. Muestras

Para simular el caso de identificación de locutor, contamos con cinco muestras recogidas personalmente por la autora de esta investigación. Todas ellas están tomadas en las mismas condiciones, para así poder estudiar todos sus rasgos de la manera más fiable posible. Los informantes de estas muestras responden a dos preguntas principales, que son:

a) *¿Cuál es tu carrera?*

b) *¿Por qué te decidiste por ella?*

Lo que se ha intentado al realizar estas preguntas es que, como hemos dicho anteriormente, la primera tuviera una respuesta breve y concisa, y en la segunda tengan libertad para expresar lo que piensan y así observar la mayor cantidad de rasgos posibles de manera espontánea. Todas las muestras tomadas rondan entre los 50 segundos y el minuto de duración.

Muestra 1: Tomada en la sala de trabajo 1 de la Facultad de Filosofía y Letras de Cádiz. Ha sido grabada a través del programa Praat, que describiremos más adelante. Se ha utilizado un ordenador personal portátil Toshiba Satélite L750/L755 y unos cascos con micrófono integrado modelo GX Gaming. El informante está en una media de edad entre 23 – 30 años. Su duración es de 50''.

Muestra 2: Tomada en la sala de trabajo 1 de la Facultad de Filosofía y Letras de Cádiz. Ha sido grabada a través del programa Praat, que describiremos más adelante. Se ha utilizado un ordenador personal portátil Toshiba Satélite L750/L755 y unos cascos con micrófono integrado modelo GX Gaming. El informante está en una media de edad entre 23 – 30 años. Su duración es de 1'05''.

Muestra 3: Tomada en la sala de trabajo 1 de la Facultad de Filosofía y Letras de Cádiz. Ha sido grabada a través del programa Praat, que describiremos más adelante. Se ha utilizado un ordenador personal portátil Toshiba Satélite L750/L755 y unos cascos con micrófono integrado modelo GX Gaming. El informante está en una media de edad entre 23 – 30 años. Su duración es de 53''.

Muestra 4: Tomada en la sala de trabajo 1 de la Facultad de Filosofía y Letras de Cádiz. Ha sido grabada a través del programa Praat, que describiremos más adelante. Se ha utilizado un ordenador personal portátil Toshiba Satélite L750/755 y unos cascos con micrófono integrado modelo GX Gaming. El informante está en una media de edad entre 23 – 30 años. Su duración es de 60”.

Muestra 5: Tomada en la sala de trabajo 1 de la Facultad de Filosofía y Letras de Cádiz. Ha sido grabada a través del programa Praat, que describiremos más adelante. Se ha utilizado un ordenador personal portátil Toshiba Satélite L750/L755 y unos cascos con micrófono integrado modelo GX Gaming. El informante está en una media de edad entre 23 – 30 años. Su duración es de 59”.

7. Estudio

7.1. Fundamentos teóricos y prácticos

En los próximos apartados se van a desarrollar los conocimientos necesarios que debemos tener sobre fonética acústica y articulatoria. Debemos conocerlos para poder realizar un estudio completo, como el que hemos llevado a cabo.

En el primer epígrafe, conoceremos todas las partes del aparato fonador y los tipos de sonidos con los que contamos en el español. Explicaremos también todos los conceptos básicos sobre el sonido, la voz, el timbre, la intensidad, el tono, etc. Luego, en el segundo epígrafe, conoceremos las distintas herramientas que utilizaremos para realizar los diferentes tipos de análisis. Para realizar estos tipos de análisis, Carlos Delgado planteaba, en su tesis (2001), un método combinado. Este método abarca un análisis acústico con un análisis articulatorio, finalizando con un pequeño análisis lingüístico. Nosotros tomaremos esto como base, pero solo realizaremos un análisis perceptivo, otro articulatorio, otro acústicos y, por último, un análisis lingüísticos.

Para Carlos Delgado (2004a), las aproximaciones clásicas expresan el resultado mediante grado de similitud (consistencia entre muestras). En muchas situaciones, quien lleva a cabo un acto delictivo de manera intencionada o planificada procura eliminar o modificar cualquier evidencia. Por lo tanto, no siempre se nos van a presentar situaciones ideales que nos permita llevar a cabo un análisis forense en condiciones óptimas. Por ello, queda justificado el empleo de escala de certeza propuesta por este autor, para expresar los resultados en este estudio, cuyos niveles serían los siguientes:

- Nivel de identificación
- Nivel de alta probabilidad
- Nivel medio – alto de probabilidad
- Nivel inconclusivo
- Nivel medio – bajo de probabilidad
- Nivel de baja probabilidad
- Nivel de eliminación

En esta investigación vamos a emplear este análisis por escalas para analizar los rasgos articulatorios de las muestras, los de registro y los idiolectales. Los utilizaremos en el apartado de dictamen, una vez que hayamos realizado los estudios oportunos. Tomando como base la escala propuesta por Carlos Delgado (2004), nosotros hemos establecido las siguientes escalas, teniendo en cuenta el número de rasgos en los que coincidan los de la muestra dubitada con los de las muestras indubitadas.

a. Para los 13 rasgos articulatorios que hemos analizado, utilizaremos la siguiente escala:

- 13 - Nivel de identificación
- 12 – Nivel de alta probabilidad
- 8-11 – Nivel medio – alto de probabilidad
- 5-7 – Nivel inconclusivo
- 3-4 - Nivel medio – bajo de probabilidad
- 1-2 – Nivel de baja probabilidad
- 0 – nivel de eliminación

b. Para los 11 rasgos idiolectales y de registro, utilizaremos la siguiente escala:

- 11 – Nivel de identificación
- 9-10 – Nivel de alta probabilidad
- 7-8 – Nivel medio – alto de probabilidad
- 5-6 – Nivel inconclusivo
- 3-4 - Nivel de medio – bajo de probabilidad
- 1-2 – Nivel de baja probabilidad
- 0 – Nivel de eliminación

Por otra parte, hemos establecido otra escala para analizar los resultados que nos proporcionan los rasgos acústicos. Esta se la vamos a aplicar a los resultados de la distancia euclídea que realizaremos más adelante, partiendo de las medidas del Pitch de las vocales analizadas. La escala establecida es la siguiente:

- 0 – Nivel de identificación
- 1-29 – Nivel de alta probabilidad
- 30-49 – Nivel medio – alto de probabilidad
- 50-99 – Nivel inconclusivo
- 100-149 – Nivel medio – bajo de probabilidad
- 150-200 – Nivel de baja probabilidad
- +200 – Nivel de eliminación

Vamos a realizar también esta probabilidad de similitud con los resultados que nos proporcionará el programa informático CopyCatch sobre el vocabulario común que existe entre la muestra dubitada y las muestras indubitadas. La escala que hemos establecido para esta observación de datos es la siguiente:

- 60-100 – Nivel de identificación
- 50-59 – Nivel de alta probabilidad
- 30-49 – Nivel medio – alto de probabilidad
- 20-29 – Nivel inconclusivo
- 10-19 – Nivel de baja probabilidad
- 0 – Nivel de eliminación

Debemos realizar también esta probabilidad de similitud con los resultados que nos proporcionará el análisis con Signature, sobre la deferencia del Chi-cuadrado de las letras y las palabras utilizadas en las muestras. Para llevar a cabo esta comparación, hemos establecido la siguiente escala:

- 1 – Nivel de identificación
- 0.5-0.9 – Nivel de alta probabilidad
- 0.06-0.4 – Nivel medio – alto de probabilidad
- 0.05 – Nivel inconclusivo
- 0.04 – Nivel medio – bajo de probabilidad
- 0.03 – Nivel de baja probabilidad
- 0-0.002 – Nivel de eliminación

7.1.1. Teóricos

Para llevar a cabo la parte del análisis articulatorio, debemos conocer qué órganos componen el aparato fonador. El cuerpo no tiene un aparato específico para la producción de sonidos, sino que aprovecha de ello los órganos que tiene otra función vital como son la respiración y la deglución. Este conjunto de órganos, al que denominamos aparato fonador, se pueden clasificar en tres grandes grupos (Quilis, A. 1999):

- a) Cavidades infraglóticas (órgano respiratorio). Estas cavidades están formadas por los pulmones, los bronquios y la tráquea. La frecuencia respiratoria es el número de veces que respiramos por minuto y es inversamente proporcional al tamaño corporal. Un recién nacido cuenta con unas 60 veces por minuto. Un niño de tres años entre 25 – 30 veces. Un adulto hombre entre 16 – 18 y una mujer entre 18 – 20 veces. Esto varía dependiendo del estado de ánimo y la actividad física. (Quilis, A. 1999).
- b) Cavidad laríngea (órgano fonador). Formada por la laringe, que a su vez se divide en tres partes. La primera es el cricoides, que es la base. La segunda es el tiroides, que se une al anterior por medio de dos conductos. Está abierto por la parte superior y la inferior. La tercera y última parte son los aritenoides, situados sobre el cricoides y movido por un sistema de músculos. Dentro de la laringe se encuentran las cuerdas vocales, que son tendones situados horizontalmente y que constituyen una parte principal del aparato fonador. Miden aproximadamente 17 a 24 mm en los hombres, de 14 a 17 en las mujeres, y son mucho más pequeñas en los niños. El paso que queda entre estas cuerdas es la denominada glotis, que se abre y cierra dependiendo de los músculos de los aritenoides. La acción de las cuerdas vocales produce una de las grandes clasificaciones de los sonidos, como es la distinción que se hace entre sonidos sonoros y sordos, que dependen de la vibración de estas. Si vibran decimos que son sonoros y si no lo hacen concluimos que son sonidos sordos. (Quilis, A. 1999).

La fase activa de la formación del tono es la tensión intencional de los músculos de la laringe contra la cual la presión del aire ejerce una acción antagonista. El sonido es un efecto conjuntado de la presión infraglótica, que es la presión del aire por debajo de la glotis, y la tensión de las cuerdas vocales. Cuanto más fuerte es la tensión, más alta es la frecuencia. (Malmerg, 1974, 115. Cit. Sánchez Ríos, J.A (2002)).

- c) Cavidades supraglóticas. Al pasar el aire por la laringe y la faringe, se realiza otra de las divisiones principales del sonido, dependiendo de la acción del velo del paladar. Si está adherido a la pared faríngea, se producen sonidos articulados orales y, por el contrario, si desciende, se articulan sonidos consonánticos nasales. Cuenta con la cavidad bucal. Esta está dividida en órgano pasivo (maxilar superior), compuesto por los alveolos y el paladar. De este último podemos diferenciar dos partes. Por un lado el paladar duro, compuesto por el prepaladar, el mediopaladar y el postpaladar, y por otro lado el paladar blando, o velo del paladar, ya mencionado anteriormente. Existe la zona prevelar y la zona postvelar. Y por órganos activos, que permiten controlar el movimiento. Abarca el maxilar inferior, la lengua, en la que se diferencian varias partes, la punta o ápice, el dorso (predorso, mediodorso y postdorso) y la raíz, los dientes, y los labios. (Quilis, A. 1999).

El sonido es un movimiento de vibración longitudinal perceptible por el oído. Las características acústicas son cada uno de los parámetros físicos medibles que tiene el sonido. Estos son, por un lado la amplitud, que es la intensidad, y nos hace diferenciar un sonido débil de uno fuerte. Se mide en decibelios (db). Por otro lado la frecuencia, por la que podemos diferenciar, debido a la cantidad de vibraciones que se generan en una unidad de tiempo, un sonido grave de uno agudo. Esta diferenciación se mide en Hercios (Hz) y se realiza concretamente por su altura. Y por último el tiempo, que tiene como característica la duración. Gracias a esta característica podemos diferenciar un sonido corto de uno largo. Esta diferencia se expresa en milisegundos (ml) o segundos (seg). (Frías Conde, X. 2001).

Todas estas características del aparato fonador nos ayudan a clasificar, dependiendo de cómo se realicen, los sonidos. La primera diferenciación que existe es la que hacemos entre vocales, que son los sonidos en los que no se interponen obstáculos a la salida del aire, y las consonantes, donde el hablante obstaculiza de alguna manera la salida del aire. (Frías Conde, X. 2001). Debemos realizar una clasificación, compuesta de cuatro puntos, para las consonantes. Para explicarlas, nos hemos apoyado en el artículo de Xavier Frías Conde sobre *la Introducción a la fonética y fonología del español* (2001).

Esta clasificación es la siguiente (Frías Conde, X. 2001):

- I. Por la acción de las cuerdas vocales. Podemos distinguir entre sordos y sonoros.
- II. Por la acción del velo del paladar. Distinguimos entre orales y nasales.
- III. Por el modo de articulación. Se produce por la modificación que el grado de abertura o cierre de los órganos articulatorios produce en la corriente de aire fonador. Dependiendo del modo de articulación, debemos distinguir que las consonantes sean:
 - a) Oclusivas. En la que se produce una explosión. Ejemplo: /p/. Se caracteriza por la existencia de una barra de explosión. El espacio entre la barra de explosión y el sonido siguiente se conoce como Voice Onset Time (VOT). En el momento de la explosión tiene valor cero, y para las oclusivas sordas, adquiere valores positivos, ya que el sonido siguiente se inicia con posterioridad, pero en el de las sonoras es negativa, porque la vocal se inicia ya con la sonoridad de la oclusiva.
 - b) Fricativas. Se producen por un roce, una fricción. Ejemplo: /f/.
 - c) Africadas. Se producen por la combinación de las oclusivas y las fricativas. Ejemplo: /tS/.
 - d) Nasales. En este caso se producen cuando parte del aire sale por la nariz. Ejemplo: /m/.
 - e) Vibrantes. Se producen cuando existe vibración de la lengua. Ejemplo: /r/.
 - f) Líquidas / laterales. En este caso la lengua hace que el aire salga por los laterales de la boca. Ejemplo: /l/.

IV. Por el punto de articulación. El lugar de articulación es el punto en el que un órgano activo, como puede ser la lengua o los labios, entra en contacto o se aproxima a un órgano pasivo para producir un mayor o menor cierre del canal bucal. Según el punto de articulación debemos distinguir consonantes entre:

a) Bilabiales. Cuenta con la participación de los dos labios. Ejemplo: /b/.

b) Labiodentales. Contacto del labio inferior con los dientes superiores.

Ejemplo: /f/.

c) Interdentales. Se dan cuando la lengua se introduce entre los dientes.

Ejemplo: /θ/ (este signo es el que representa el sonido <z>, de “zapato”).

d) Dentales. Existe contacto entre la lengua y la parte trasera de los dientes superiores. Ejemplo: /d/.

e) Alveolares. Se producen cuando la lengua toca los alveolos superiores.

Ejemplo: /s/.

f) Palatales. Se dan cuando la lengua toca el paladar. Ejemplo /tS/. (Este signo es la representación de nuestra <ch> ortográfica.

g) Velares. Existe cuando hay contacto de la lengua con el velo del paladar.

Ejemplo: /k/.

h) Uvulares. Se articula con la parte posterior de la lengua colocada más atrás que las velares. Ejemplo: /q/

i) Faringales. La raíz de la lengua se coloca cerca de la pared faríngea.

Ejemplo: /ħ/

j) Glotales. Se producen por el cierre de la glotis. Ejemplo: /h/.

Aquí podemos observar una tabla que nos ayudará como apoyo a estas clasificaciones antes mencionadas. Los sonidos son representados por el Alfabeto de Fonética Internacional (IPA), que es un sistema de representación de sonidos de alcance internacional.

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b		t d			ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s ʃ	ʂ ʐ	ʐ ʑ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

Este alfabeto es el que se recomienda utilizar para las transcripciones, por ello este es el sistema que nos ayudarán posteriormente para la realización de las transcripciones de las muestras que debemos analizar.

En cuanto a las vocales debemos saber que en castellano solo existen cinco, y que se clasifican de la siguiente manera: (Frías Conde, X. 2001).

	iniciales	centrales	finales
cerradas	/i/		/u/
medias		/e/	/o/
abiertas		/a/	

Según Xavier Frías (2001), la /i/ es una vocal inicial cerrada. La /u/ es final cerrada. La /e/ es media inicial – central y la /o/ es media final – central, mientras que la /a/ es central abierta. Las vocales deben clasificarse entre:

- Fuertes, que son /a/, /e/ y /o/.
- Débiles, que son /u/ e /i/.

Este autor, en 2001, también nos avisa de que debemos tener en cuenta los diptongos, que son los encuentros entre una vocal débil y una fuerte en una misma sílaba, formando una sola vocal. A efectos fónicos y ortográficos, son vocales simples. Los triptongos por su parte se forman cuando hay cúmulo de tres vocales, siempre que la primera y la tercera sean débiles.

Cuando producimos sonidos estamos poniendo en funcionamiento un complejo mecanismo en nuestro cuerpo que implica, por un lado, a numerosos órganos y, por otro, a una determinada coordinación temporal y espacial entre ellos para que el resultado sea satisfactorio. (Fernández Planas, Ana M., 2005).

Por otro lado, para llevar a cabo el análisis acústico, debemos saber que las vocales cuentan con unas características que debemos medir, porque dependiendo de la medida de estos, podremos deducir de qué vocal se trata dentro de un espectrograma. Este elemento son los formantes. Estos son los conjuntos de frecuencias que se ven reforzadas como consecuencia del fenómeno de la resonancia. Es una gama de frecuencias, un conjunto de armónicos que presentan más intensidad. (Frías Conde, X. 2001).

La extracción de los formantes es un objetivo necesario, ya que cuenta con una gran importancia para conseguir una correcta interpretación de los espectros de voz. Una vez conocidos los formantes, es necesario conocer sus peculiaridades, que varían dependiendo del hablante, los distintos contextos del sonido y la entonación de cada frase. Es necesario saber que solo las vocales, las laterales y las nasales cuentan con formantes, y debemos dar a conocer el sistema de medidas por el que nos vamos a regir para nuestro análisis. Los formantes que debemos medir son F1, que es el formante de frecuencia más bajo y F2, que es el segundo. No existen valores absolutos para las vocales, dependen de la persona, pero ese abanico de valores por los que ronda es conocido como campo de dispersión. (Bobadilla, J., Gómez, P. y Bernal, J., 1999)

Para analizar la medición de F1 y F2, vamos a tomar como referencia los valores de los formantes vocálicos de Martínez Celdrán (1998). Estos deben ser diferentes entre la voz masculina y la femenina, pero los valores de cada sonido deben rondar los siguientes:

VOCALES	F1	F2
A	700	1500
E	500	1800
I	300	2000
O	500	1800
U	300	700

Junto con el análisis de los formantes, debemos atender también al acento fonético. Este acento es la mayor fuerza en la expulsión de aire, duración y una ligera del tono fundamental. La entonación, que depende de la longitud, el grosor y la tensión que se les da a las cuerdas vocales. Y la duración, que constituye el tiempo de prolongación de los sonidos o sílabas, y dependen de la caracterización idiolectal del hablante. (Frías Conde, X. 2001).

7.1.2. Herramientas

Para realizar los análisis, utilizaremos los siguientes programas:

- a) *Praat*, que es una herramienta para el análisis fonético del habla. Realiza un análisis acústico, con síntesis articulatoria, procesamiento estadístico de los datos, edición y manipulación de señales de audio. Fue desarrollada por Paul Boersma y David Weenink en el Instituto de Ciencias Fonéticas de la Universidad de Ámsterdam, a partir de 1992. (BOERSMA, P. y D. WEENINK (2008), cit Román, D. (2008)). Es un software totalmente gratuito, que se puede obtener en la página web www.praat.org.

Domingo Román (2008), del Laboratorio de Fonética de la Facultad de Letras de la Pontificia Universidad Católica de Chile, ha elaborado un manual básico para la utilización de esta herramienta. *Praat* es un programa de libre distribución, de código abierto, multiplataforma, gratuito y es actualizado con mucha frecuencia.

Al ser un programa especializado es necesario tener una serie de conocimientos teóricos, para llevar a cabo una buena interpretación de los datos obtenidos. Debemos saber que esta herramienta nos proporciona la imagen del sonido en espectrogramas. Esta gráfica sitúa en el eje de abscisas el tiempo y en el eje de ordenadas, la frecuencia (Hz). Refleja el sonido en su evolución frecuencial a lo largo del tiempo. Pueden ser de dos tipos (Domingo Román, 2008):

1. De banda ancha, que no permite ver los armónicos sino sus concentraciones a ciertas frecuencias. Son de 150 Hz.
2. De banda estrecha, donde se puede ver los armónicos. Son de menos de 15 Hz.

Según nos explica Xavier Frías (2001), en el *praat* se pueden observar los formantes como líneas formadas por puntos rojos superpuestos. Algunas veces podemos ver líneas de puntos que no llegan a ser formantes, sino que son resonancias propias de una emisión de una persona.

b) *Transcriber*.⁴ Esta herramienta nos ayuda en el proceso de transcripción, con el que podremos realizar un análisis perceptivo. Es un programa de código abierto. Ayuda a organizar la información en una pantalla segmentada en dos partes. En la parte superior, la persona que realiza la transcripción puede incorporar a los hablantes indicando los puntos del debate o entrevista, con la ayuda de iconos. Todas las transcripciones pueden ser exportadas a diferentes tipos de archivos para llevar a cabo su tratamiento. (Sánchez Gómez, M.C. y Revuelta Domínguez F.I. 2005).

En esta herramienta debemos siempre identificar la forma de transcribir a la que nos estamos adaptando, es decir, indicar si hemos seguido las normas de la gramática del idioma o si por el contrario nos hemos ceñido a representar todos los sonidos tal y como los escuchamos. Para ellos debemos realizar una lista de las referencias de cada uno de los símbolos que hemos utilizado.

Existen tres formas diferentes de transcribir. Podemos reflejar solo aquellos sonidos que difieran de lo normal, reflejar solo las principales características de los sonidos o reflejar cualquier característica de los sonidos. Todo ello tomando los elementos del Alfabeto fonético internacional (IPA). (Sánchez Gómez, M.C. y Revuelta Domínguez F.I. 2005). Como ejemplo:

El trabajo nos alivia el dolor.
 TR.A: [el tra'βaxo nos a'liβja el do'lor]]
 TR.E: [e] tra'βaxo↑ nos a'liβja e] do'lor↓]].

Las transcripciones realizadas serán realizadas reflejando solo los rasgos de aquellos que difieran de lo normal. Las guardaremos en formato de bloc de notas (*.txt) y serán presentadas en los anexos de este trabajo.

- c) *CopyCatch* (<http://www.copycatch.free-serve.co.uk/vocalyse.htm>) nos será de mucha ayuda para identificar el número de palabras compartidas en las distintas muestras. Es una herramienta para el procesamiento de textos. Este muestra únicamente las palabras y oraciones compartidas por dos archivos determinados. Se creó con el objetivo de ofrecer una visión rápida del nivel de similitud entre varios pares de textos, concretamente redacciones de clase sobre un mismo tema.

Los resultados de salida se han optimizado para que incluyan tanto los pares de segmentos, como de oraciones, y cualquier par textual pueda resaltarse para mostrar cómo aparece el léxico compartido en el texto: en modo de archivos del todo emparejados, o simplemente en modo de oraciones con material compartido, y toda una serie de combinaciones entre sí. (Turell. M.T. 2005).

d) *Signature*. Este programa te permite contrastar la independencia de los datos de dos textos mediante el test chi – cuadrado. El estándar científico considera que los elementos son estadísticamente diferentes si están al 0,5%, es decir, 0.05 o menos. Esto significa que si deberíamos considerar los textos como provenientes de autores diferentes. Si no lo hallamos deberíamos pensar que ambos tienen el mismo autor. Contrasta “Word length”, “sentence length”, “paragraph length”, “letters” y “punctuation” de los textos dubitados con indubitados usando este test.

e) *Microsoft Excel* (<https://office.live.com/start/Excel.aspx?omkt=es-ES>). Para la recopilación y comparación de resultados. Como bien nos explica Agustín Cernuda (2001), este programa se caracteriza por la utilización de las hojas de cálculo que nos ayudan a la inclusión de tablas en los documentos que se manejan. Recogeremos en estas tablas todos los datos con el fin de tenerlos organizados para una posterior comparación. Gracias a este programa podemos también realizar gráficas representativas de los datos que creamos oportunos.

f) *Statdisk* (<http://www.statdisk.org/>). Para ayudarnos a realizar comparaciones entre las medias de los valores de las vocales de los hablantes, con el fin de obtener un P-value que nos ayude a comprobar o rechazar las hipótesis. Con estos resultados también podemos saber si han salido esos resultados porque los elementos son iguales o si son totalmente diferentes. Nos permite realizar gran cantidad de test de hipótesis. Nosotros vamos a utilizar exclusivamente uno de los que se encuentra dentro del apartado *Analysis*, que se encuentra en la parte superior de la barra de tareas del programa. Dentro de este desplegable que nos aparece debemos seleccionar *Hypothesis testing*, y dentro del siguiente menú que se nos despliega, elegimos el test *Mean - two independent samples*. Este test es el que se utiliza para comparar dos proporciones tomadas de dos muestras. Si los valores que nos aporten estos test en P-Value son < 0.05 , determinaremos que las muestras presuntamente son de distintos hablantes, mientras que si estos valores son > 0.05 , determinaremos que las muestras son, presuntamente, de la misma persona.

7.1.3. Análisis

Hemos realizado esta investigación dividiéndola en cuatro partes principales. Cada una de ellas recoge los distintos análisis que vamos a realizar sobre las muestras recopiladas.

En primer lugar vamos a llevar a cabo un análisis perceptivo, en el que, sacaremos conclusiones sobre los rasgos comunes y los distintivos de cada individuo, a partir de las escuchas. Debemos escuchar las muestras repetidas veces y con detenimiento, para sacar los rasgos que seamos capaces de distinguir en cada uno de los casos. Para este tipo de análisis no necesitaremos ningún tipo de programa informático. Nos bastará con unos buenos cascos para lograr una escucha óptima.

Seguiremos con un análisis articulatorio, donde contaremos con una serie de programas informáticos que nos ayudarán a realizar una transcripción, para observar rasgos de pronunciación particular, los rasgos idiolectales de los individuos que debemos analizar. Esta transcripción será realizada siguiendo el Alfabeto Fonético Internacional (IPA), del cual tenemos una tabla de referencia en el epígrafe anterior. El Transcriber será el soporte informático que nos ayudará en este caso.

Luego, realizaremos un análisis acústico. Nos centraremos en la delimitación y medición de las vocales, en especial las llamadas vocales fuertes, como son [a], [o] y [u]. Para realizar este análisis nos ayudaremos de un programa informático llamado Praat. Una vez medidas estas vocales, tomaremos el valor medio del Pitch de cada una de ellas y calcularemos la distancia euclídea, lo que nos ayudará a observar qué muestra se acerca más a la dubitada en la pronunciación de dichos sonidos. Si el análisis realizado no nos proporciona unos resultados totalmente fiables, buscaremos alguna palabra que aparezca en todas las muestras y le realizaremos el mismo análisis que a las vocales, es decir, mediremos la palabra, tomaremos el valor medio del Pitch y calcularemos su distancia. Además de estos análisis realizados, haremos un Test de Hipótesis mediante el programa informático Statdisk, para ver los resultados que nos proporciona el P-Value de la comparación de la muestra dubitada con cada una de las muestras indubitadas y ver si alguna puede parecerse o no.

Por último llevaremos a cabo un análisis lingüístico, en el que observaremos la frecuencia de palabras iguales, mediante el programa CopyCatch y realizaremos también el estudio del chi-cuadrado de todas las muestras para hacer una comparativa de la similitud entre muestras. En este caso utilizaremos el programa informático Signature.

Todos los programas informáticos mencionados ya han sido explicados detalladamente en el epígrafe anterior.

7.2. Cotejo

7.2.1. Análisis articulatorio

Para comenzar con el análisis en profundidad, lo primero que vamos a realizar es un análisis perceptivo, como proponía Carlos Delgado (2001), a través de un análisis auditivo y una transcripción realizada con el programa antes mencionado *Transcriber*. Según Carlos Delgado (2004), el proceso perceptivo es más complejo que una simple audición. Es un proceso que necesita experiencia y entrenamiento del experto, para poder discriminar y contextualizar algunas de las características de los eventos hablados de un locutor determinado, como son por ejemplo las peculiaridades del sociolecto, los recursos retóricos, la ubicación de la base de la articulación, el timbre, etc.

El conocimiento sobre geografía dialectal, lexicografía y sociolingüística puede ayudar en la investigación de un crimen. (Jiménez, M. et al, 2012). No se puede hablar de lengua sin hablar de sociedad sin tener en cuenta que la lengua es un hecho social y que todos los demás hechos sociales se vinculan a ella. (D.Cassany, 2002). Una vez realizadas las transcripciones de todas las muestras, vamos a compararlas atendiendo a los rasgos idiolectales que presente el individuo y el registro que presente. Los rasgos idiolectales son los que debemos tener en cuenta fijándonos en la forma de hablar del individuo. Cada persona tiene una forma propia y característica de hablar. No todos los hablantes usan los signos y las reglas de combinación de la misma manera al hablar. La selección de las variantes se relacionan con las diferencias sociales y culturales entre individuos, como las clases sociales o los grupos étnicos, etc. (Barrios y Coll, 1994 cit. Casales,F. 2005).

Debemos realizar una distinción entre lengua, dialecto e idiolecto. Consideramos dialecto como diferenciación de un grupo. Esta depende del lugar donde se desarrolla la lengua y el lugar en que residen. A esta la denominamos variedad diatópica. Llamamos variedad diastrática al nivel de instrucción del individuo y variedad diafásica a la situación comunicativa en la que nos encontramos. Por su parte, la lengua cuenta con una evolución a lo largo de la historia, lo que nos permite realizar un estudio sobre la variación diacrónica. (Barrios y Coll, 1994 cit. Casales, F. 2005).

Como podemos encontrar en el Diccionario de la Real Academia Española (DRAE), el idiolecto es un conjunto de rasgos propios de la forma de expresarse de un individuo. En el trabajo de Roberto Mayoral Asensio, de la Universidad de Granada (UGR), sobre La traducción de la variación lingüística, podemos ver cómo nos habla del idiolecto desde el concepto de Halliday y otros (1964). Para estos autores el idiolecto es el “estilo individual” y se incluye dentro de la categoría de registros, variedad de uso. Muchos autores han colocado el idiolecto en diferentes categorías de la lengua. Roberto Mayoral Asensio, en su tesis doctoral de la Universidad de Granada (1998), nos habla sobre el idiolecto, y nos advierte que no podemos considerarlo como inmóvil, ya que no siempre es el mismo aunque se trate de la misma persona. Una persona puede modificar su forma de hablar a lo largo de su vida. Language and National Origin Group (2004) advierte que el análisis lingüístico no puede usarse con total fiabilidad para determinar el origen nacional, ya que el origen, la nacionalidad, etc, son conceptos burocráticos. Este análisis lingüístico puede usarse para extraer conclusiones razonables, pero no siempre el país de socialización del hablante es el de su origen. Además, el país de socialización solo puede estimarse cualitativamente, es decir, posible, probable, muy probable o muy poco probable.

Nosotros en este trabajo nos vamos a centrar, sobre todo, en los rasgos del andaluz, ya que nuestras muestras cuentan con muchos de estos rasgos.

El andaluz o español meridional es una variedad o modalidad lingüística que ocupa el sur de la Península Ibérica, desde el Mediterráneo oriental almeriense hasta el río Guadiana, frontera político administrativa entre España y Portugal, cuyas aguas bañan tierras portuguesas del Algarve y andaluzas de Huelva. (Modéjar, J. 1991)

De estos rasgos, debemos destacar los que se consideran más significativos, los más característicos, como son el seseo – ceceo, aspiración de /s/, aspiración de /x/, el yeísmo, el tratamiento de /l/ y /r/, la fricativización de /ç/, la /d/ final ante pausa, el “heheo” o el tratamiento de /d/ intervocálica. (Jiménez Fernández, R. 1999).

- *Seseo – ceceo*: es uno de los rasgos más identificativos del andaluz. Esto sirvió para trazar fronteras lingüísticas, en 1933. Esto fue realizado en un estudio del fonetista Tomás Navarro, que no solo estableció la extensión geográfica de las zonas seseantes y ceceantes, sino que, además, analizó por primera vez los diferentes tipos de /s/ y /θ/ del andaluz. Existen tres modelos lingüísticos posibles, como son, el seseo, que se caracteriza por la reducción de fonemas /θ/ y /s/ a uno solo realizado fonéticamente como [s], el ceceo, en el que se realiza una reducción de fonemas /θ/ y /s/ a uno solo realizado como [θ] y la distinción, que consiste en la diferenciación de las realizaciones de ambos fonemas. (Jiménez Fernández, R. 1999).

- *Aspiración de /s/*: este es otro de los rasgos más característicos de la fonética andaluza. Consiste en la aspiración y pérdida de /s/ cuando se encuentra en posición final de sílaba ante otra consonante o como final absoluta. Ejemplo: [loh – ómbreh]. (Jiménez Fernández, R. 1999).

- *Aspiración de /x/*: es el fonema velar fricativo sordo. Tiene diversas realizaciones fonéticas. Dos de las más importantes son las siguientes. Por un lado la realización como velar sorda [x], que se articula con el postdorso de la lengua contra el paladar, haciendo que el aire fonador circule por un espacio muy estrecho. Se da en la península, desde el centro hasta el norte. Como ejemplo de esta representación tenemos jirafa [xiráfa]. Por otra parte tenemos la representación que nos interesa en la mayoría de las muestras que tenemos que analizar. Este caso es una representación aspirada [h]. Se considera faríngea si se articula con la raíz de la lengua contra la pared faríngea o laríngea, si se

articula por medio de un estrechamiento de las cuerdas vocales. Como por ejemplo: ojo [óho]. En Andalucía puede producirse como una suave aspiración farígea sorda [h] o se puede articular como sonora [h]. (Jiménez Fernández, R. 1999).

- El *yeísmo*: lo entendemos como la realización de la misma pronunciación de los sonidos que representamos gráficamente como 'y' y 'll'. El yeísmo se produce por la confusión de los dos sonidos, en definitiva, realizar la articulación como 'y' de toda 'll'. (Narbona,A., Cano, R y Morillo, R. 1998).

- *El tratamiento de /l/ y /r/*: este fenómeno consiste en la pérdida del carácter distintivo de la vibrante alveolar /r/ y la lateral alveolar /l/. El sonido se suele realizar como una articulación relajada, realizada como [r], pero también como [l], o sencillamente como aspirada [h]. Por ejemplo: soldado [sordáo] o alto [árto]. (Jiménez Fernández, R. 1999).

- *Fricatización de /ç/*: consiste en la articulación relajada del fonema africado, convirtiéndose así en fricativa, más próxima a la 'sh' inglesa o 'ch' francesa. En Andalucía presenta diversas realizaciones fonéticas tanto por el modo como por el lugar de articulación. Las dos variantes más interesantes son, por un lado la articulación africada, que consta de un momento oclusivo seguido de otro fricativo y por otro lado la articulación fricativa. Esta conversión de africada a fricativa, se debe a la desaparición del sonido oclusivo por una relajación articulatoria. (Jiménez Fernández, R. 1999).

- */d/ final ante pausa*: desaparece en casi todo el español meridional y muy intensamente en Andalucía. Ejemplo: pared [paré] o red [ré]. Estas son las pronunciaciones normales y prácticamente uniformes en Andalucía. El único resto de esta consonante que queda en la pronunciación de la palabra es una leve abertura de la vocal final. (Narbona,A., Cano, R y Morillo, R. 1998).

- *El “heheo”*: consiste en una aspiración de la /s/ o /z/ inicial de palabra o sílaba. (Narbona,A., Cano, R y Morillo, R. 1998).

- *El tratamiento de /d/ intervocálica*: su pérdida se puede producir entre cualquier vocal. Podemos estudiarla entre vocales de la misma naturaleza, como por ejemplo, cansada [cansá] o entre vocales de diferente naturaleza, como por ejemplo cantado [cantao]. (Jiménez Fernández, R. 1999).

También debemos tener en cuenta los vulgarismos. Existen una serie de rasgos lingüísticos que son propios de los hablantes de menor cultura, aunque en este caso concreto no tienen por qué ser por esta razón. Se puede deber a rasgos lingüísticos de una conversación relajada, informal, en un registro coloquial. No pueden considerarse rasgos dialectales, ya que no existe una delimitación geográfica en la que aparecen. (Narbona,A., Cano, R y Morillo, R. 1998). Dentro de estos vulgarismos encontramos la pérdida de /d/ intervocálica, mencionada anteriormente. Socialmente está considerada como tal y, por tanto, no se aconseja en registros cultos, quedando relegada a hablas familiares o coloquiales. La Real Academia toleraba, en 1999, la relajación y pérdida de –d- en los participios en –ado. Cuando nos fijamos en el registro, en el caso de las muestras estudiadas, es un registro coloquial, debemos atender al número de pausas que realiza el individuo, a la velocidad a la que habla, a balbuceos, a reformulaciones, etc. (Jiménez Fernández, R. 1999).

Una vez realizado el estudio mencionado anteriormente, vamos a hacer una tabla comparativa con los rasgos más relevantes y significativos que hemos observado. No pondremos todos los casos que se representan en las muestras, solo pondremos algunos ejemplos de cada uno. Cabe destacar que solo señalaremos la transcripción del rasgo concreto que estamos observando.

Rasgo fonético - articulatorio	Muestra Dubitada	Muestra indubitada 1	Muestra indubitada 2	Muestra indubitada 3	Muestra indubitada 4
Aspiración de [s] (sonido oral, sordo, linguo – dental, fricativo) en posición implosiva. Esta Aspiración [h] es un sonido oral, sonoro, glotal.	No se observa	Lingüística: Lingüí[h]tica Estuve: E[h]tuve Estaba: E[h]taba Hasta: Ha[h]ta Listas: Li[h]ta	Lingüística: Lingüí[h]tica Esta: E[h]ta Gustan: Gu[h]tan Gustaría: Gu[h]taria	Lingüística: Lingüí[h]tica Después: de[h]pué[h]	Lingüística: Lingüí[h]tica Estar: E[h]tar Esta: E[h]ta Gustado: Gu[h]tado
“Heheo”	Necesitaba: Ne[h]esitaba	No se observa.	No se observa.	No se observa.	No se observa.
Omisión de [s] final de palabra (sonido oral, sordo, linguo – dental, fricativo). Este sonido final puede aparecer en ocasiones como aspirado [h].	No se observa	Pues: Pue[h] Los: Lo[h] Dos: Do[h] Años: Año[h] Las: La[h] Carreras: Carrera[h]	Aplicadas: Aplicada[h] Los: Lo[h] Idiomas: Idioma[h]	Lenguas: Lengua[h] Las: La[h] Letras: Letra[h] Ciencias: Ciencia[h]	Lenguas: Lengua[h] Aplicadas: Aplicada[h] Es: E[h] Más: Má[h] Menos: Meno[h]
Debilitamiento de [d]/ [ð] (sonido oral, sonoro, linguodental, oclusivo): tienden a desaparecer.	No se observa	Cambiado: cambio' Toda: toa'	No se observa	No se observa	Complicado: complicao' Gustado: gustao' Encontrado: encontrao'
Glotalización de [x] (sonido oral, sordo, linguo-velar, fricativo), tiende a relajarse, asemejándose más a un sonido [ħ] (oral, sonoro, faríngeo, fricativo)	No se observa	No se observa	No se observa	Cogí: Co[ħ]í	Filología: filolo[ħ]ía Ejemplo: e[ħ]emplo
Pérdida de [r] (sonido oral, sonoro, linguo -alveolar, vibrante simple) al final de palabra.	No se observa	Entrar: entrá	No se observa	Estudiar: estudiá Mejor: mejó	Encontrar: encontrá Completar: completá
Alargamiento de vocales y consonantes al final de palabra.	Porque: [porque:] Que: [Que:] Un: [Un:]	Pues: [Pues:] Y: [i:] La: [la:] Estaba: [Estaba:] De: [De:] El: [El:] Más: [Más:]	Porque: [Porque:] Pero: [Pero:] Dedicarme: [dedicarme:] Parecido: [parecido:]	No: n[o:] Me: m[e:] Ha: h[a:] Pero: per[o:]	La: [la:] Porque: [porque:] En: [en:] Pero: [pero:] A: [a:] Me: [me:] Y: [i:] Como: [como:]

Cambio de [tʃ] a [ʃ]	No se observa	Derecho: dere[ʃ]o Chorrada: [ʃ]orrada	Mucho: mu[ʃ]o	No se observa	No se observa
Rotacismo de sonidos líquidos [l] [r]. Pronunciación de [l] (sonido oral, sonoro, linguoalveolar, líquido), cuyo resultado es el sonido [r] (sonido oral, sonoro, linguoalveolar, vibrante).	No se observa	Sueltas: sue[R]ta[h]	No se observa	No se observa.	El: e[R] Al: a[R]
Seseo, variante de [θ] (sonido oral, sordo, linguo – dental, fricativo), pronunciado como [s] (sonido oral, sordo, linguo – dental, fricativo)	No se observa	No se observa	No se observa	No se observa	Hice: hi[s]e Entonces: Enton[s]e[h]
Sinalefas	No se observa	No se observa	No se observa	Cursos superiores en: Curso'superior e[e:][h]'n	No se observa.
Pérdida de consonantes intervocálicas	No se observa	Cambiado: Cambio Toda: Toa	No se observa	Relacionadas: Relacioná[h]	Complicado: complicao Dado: dao Encontrado: encontrao
Pérdida de consonantes finales	No se observa	No se observa	No se observa	Verdad: verdá	Verdad: verdá

Todas las representaciones de los sonidos para llevar a cabo la transcripción han sido obtenidas en el Alfabeto Fonético Internacional (IPA).

Existen otras características que debemos tener en cuenta, y dependen del registro utilizado y del idiolecto de cada hablante, como ya hemos dicho anteriormente. El hablante se relaja y cometen ciertos errores, como son los vulgarismos.

Lo vamos a reflejar, como ya hemos hecho, con una tabla comparativa.

Ratios Elocutivos	Muestra Dubitada	Muestra Indubitada 1	Muestra Indubitada 2	Muestra Indubitada 3	Muestra Indubitada 4
Fluidez de producción	Normal	Normal	Normal	Normal	Normal
Pausas	No	No	No	No	Si
Simplificaciones consonánticas	No se observa	Toda: [to] Pues: [po]	No se observa	Pues: [po] Toda: [to]	Pues: [po]
Pérdida de consonantes intervocálicas	No se observa	Tendencia a omitir /d/: Cambiado: Toda: Tendencia a omitir /r/: No se observa	No se observa	No se observa	Tendencia a omitir /d/: Complicado: Gustado: Encontrado: Tendencia a omitir /r/: No se observa
Pérdida de consonantes finales	No se observa.	No se observa.	No se observa.	No se observa.	No se observa.
Repeticiones – interrupciones – autocorrecciones	Hacer... hacer lingüística. Necesitaba... era consciente que necesitaba. Decidí que... me parecía mucho.	Y... Y... Y oposité. a la... a la guardia. de... de carrera. El... me ofrecieron	No se observa.	No se observa.	El... el... el e ^[h] pe[s]ífico.
Risas	Si	Si	No	No	No
Coger aire con fuerza	Si	Si	Si	Si	No
“Eeee”	Si	No	Si	No	Si
“Ay”	Si	No	No	No	No
“No se”	Si	No	No	No	No

Muestra dubitada – Muestra indubitada 1

Coinciden en la glotalización de [x] (sonido oral, sordo, linguo – velar, fricativo), tiende a relajarse, asemejándose más a un sonido [ħ] (oral, sonoro, faríngeo, fricativo). En ambos casos no se observa este rasgo. También podemos observar en las dos muestras que se produce un alargamiento de vocales y consonantes al final de palabra. En ninguno de los dos casos se observan seseos, sinalefas ni pérdidas de consonantes finales. No cuentan, perceptivamente con grandes pausas al hablar. Por el contrario, las dos muestras tienen repeticiones, interrupciones o autocorrecciones durante su producción. Ambas coinciden en risas y en pausas para coger aire con fuerza.

Muestra dubitada – Muestra indubitada 2

Estas dos muestras coinciden en el debilitamiento de [d] / [ð] (sonido oral, sonoro, linguodental, oclusivo), que tienden a desaparecer. La coincidencia en este caso es porque no se observa este rasgo en ninguna de ellas. Tampoco se observa la glotalización de de [x] (sonido oral, sordo, linguo – velar, fricativo), tiende a relajarse, asemejándose más a un sonido [ħ] (oral, sonoro, faríngeo, fricativo) ni la pérdida de [r] (sonido oral, sonoro, linguo – alveolar, vibrante simple) al final de palabra. No se observa en ninguna de las muestras ningún rasgo de rotacismo de sonidos líquidos [l], [r] (Pronunciación de [l] (sonido oral, sonoro, linguoalveolar, líquido) cuyo resultado es el sonido [r] (sonido oral, sonoro, linguoalveolar, vibrante)). Por el contrario, ambas coinciden en el alargamiento de vocales y consonantes al final de palabra. de [x] (sonido oral, sordo, linguo – velar, fricativo), tiende a relajarse, asemejándose más a un sonido [ħ] (oral, sonoro, faríngeo, fricativo).

No se observan seseos, sinalefas ni pérdidas de consonantes intervocálicas ni finales. Tampoco se observan perceptivamente ninguna pausa grande, ni simplificaciones consonánticas. Ambos toman aire con fuerza, hacen uso de la muletilla “eee” para momentos pensativos.

Muestra dubitada – muestra indubitada 3

En ninguna de las dos muestras se observa debilitamiento de [d] / [ð] (sonido oral, sonoro, linguodental, oclusivo), que tienden a desaparecer ni el rotacismo ni cambio de [tʃ] a [ʃ].

El alargamiento de vocales y consonantes al final de palabras y el rotacismo de sonidos líquidos [l], [r] (Pronunciación de [l] (sonido oral, sonoro, linguoalveolar, líquido) cuyo resultado es el sonido [r] (sonido oral, sonoro, linguoalveolar, vibrante)) aparecen en ambas muestras. No se observan seseos ni pérdidas de consonantes intervocálicas ni finales. Además ninguna tiene grandes pausas (desde un análisis perceptivo).

En ambos casos toman aire con fuerza.

Muestra dubitada – Muestra indubitada 4

En este caso, en ninguna de las muestras se observa un cambio de [tʃ] a [ʃ]. Tampoco se observan sinalefas ni pérdidas de consonantes finales.

En ambos casos hay repeticiones, interrupciones o autocorrecciones y utilización de la muletilla “eee” para mantener un tiempo de pensamiento.

7.2.2. Análisis Acústico

Ahora nos centraremos en el análisis fonético – acústico. Como hemos explicado en los epígrafes anteriores, nos vamos a ayudar del programa Praat⁵, para tomar las medidas necesarias para dicho estudio. Analizaremos los espectrogramas de las muestras que nos ha proporcionado el programa utilizado. Tendremos en cuenta la sonoridad, la duración, la intensidad, las pausas, la velocidad ilocutiva y el ritmo.

Sometemos a análisis tanto la muestra dubitada como las cuatro muestras indubitadas. Todos los resultados tomados los vamos a recoger en una tabla, en la que se incluirán parámetros estadísticos, con los valores medios de los formantes F1 y F2 de las vocales fuertes que coincidan entre las muestras indubitadas y la muestra dubitada.

Hemos tomado un total de seis vocales. Dos [a], una vocal átona final, perteneciente a la palabra “lingüística” y otra tónica. Dos [o], un tónica perteneciente a la artículo determinado “los” y una átona, colocada a final de palabra y dos [u] átonas situada antes de nasal. Esta última selección hemos decidido hacerla así por la dificultad de encontrar vocales [u] que sean óptimas para realizar el análisis. Por ello, hemos decidido tomar dos con las mismas características, pero en ningún caso se han repetido.

También debemos tener en cuenta la medida de F0, es decir, de la frecuencia absoluta.

⁵ Praat v. 5.3.81 (julio 2014) Boersma P. Weenik D. Universidad de Amsterdam, <http://www.fon.hum.uva.nl/praat/>

- Vocal [a]

VALOR MEDIO F1 – F2. VOCAL FUERTE [A]					
	Muestra dubitada	Muestra indubitada 1	Muestra indubitada 2	Muestra indubitada 3	Muestra indubitada 4
	[a] final de “Lingüística”				
Valor medio Pitch	128.69 Hz	282.18 Hz	149.68 Hz	250.75 Hz	162.65 Hz
Valor medio F1	648.4 Hz	888.8 Hz	517.3 Hz	692.1 Hz	692.1 Hz
Valor medio F2	1544 Hz	1457 Hz	1807 Hz	2134 Hz	1763 Hz
VALOR MEDIO F1 – F2. VOCAL FUERTE [A]					
	Muestra dubitada	Muestra indubitada 1	Muestra indubitada 2	Muestra indubitada 3	Muestra indubitada 4
	“á” entre consonantes				
Valor medio Pitch	142.72 Hz	202.71 Hz	147.39 Hz	206.53 Hz	141.16 Hz
Valor medio F1	648.4 Hz	910.7 Hz	670.3 Hz	910 Hz	845.1 Hz
Valor medio F2	1282 Hz	1238 Hz	1217 Hz	1413 Hz	1435 Hz

- Vocal [o]

VALOR MEDIO F1 – F2. VOCAL FUERTE [O]					
	Muestra dubitada	Muestra indubitada 1	Muestra indubitada 2	Muestra indubitada 3	Muestra indubitada 4
	Vocal [o] en “los”				
Valor medio Pitch	175.63 Hz	227.24 Hz	121.67 Hz	246.96 Hz	143.70 Hz
Valor medio F1	539.2 Hz	673.5 Hz	542.8 Hz	594.9 Hz	604.7 Hz
Valor medio F2	1309 Hz	1252 Hz	1010 Hz	1230 Hz	1348 Hz
	Vocal [o] en final de palabra				
Valor medio Pitch	145.08 Hz	214.50 Hz	126.40 Hz	230.49 Hz	141.44 Hz
Valor medio F1	626.6 Hz	648.4 Hz	517.3 Hz	561 Hz	561 Hz
Valor medio F2	1304 Hz	1566 Hz	1326 Hz	1588 Hz	1151 Hz

- Vocal [u]

VALOR MEDIO F1 – F2. VOCAL FUERTE [U]					
	Muestra dubitada	Muestra indubitada 1	Muestra indubitada 2	Muestra indubitada 3	Muestra indubitada 4
	Vocal [u] antes de [n]				
Valor medio Pitch	153.86 Hz	241.83 Hz	143.00 Hz	216.74 Hz	142.11 Hz
Valor medio F1	517.3 Hz	692.1 Hz	495.5 Hz	473.6 Hz	539 Hz
Valor medio F2	1173 Hz	1457 Hz	1020 Hz	1195 Hz	1085 Hz
	Vocal [u] antes de nasal				
Valor medio Pitch	152.890 Hz	252.272 Hz	143.040 Hz	198.091 Hz	140.385 Hz
Valor medio F1	556.9 Hz	683.0 Hz	453.4 Hz	494.4 Hz	477.942 Hz
Valor medio F2	1057 Hz	1382 Hz	977 Hz	887 Hz	962 Hz

Ahora vamos a contrastar mediante el *test de Student* los valores del pitch, para ello vamos a usar el programa *Statdisk* que lo hace fácilmente. Mediante este programa informático vamos a realizar un análisis comparativo calculando el P-Value. Esto datos nos pueden ayudar a saber si alguno de los hablantes de la muestras indubitadas pueden ser la misma persona que el autor de la muestra dubitada. Para ello vamos a utilizar, dentro de los Hypothesis testing que nos proporciona el programa, un test denominado Mean – Two Independent Samples. Para ello debemos conocer la media del pitch, el número de pulsos y la desviación de cada una de las vocales estudiadas. Vamos a realizar una media por un lado de media del pitch y por otro lado de la desviación de todas las vocales fuertes analizadas en cada una de las muestras. Luego, vamos a sumar el número de pulsos de todas las vocales en cada una de las muestras, ya que, para llevar a cabo este tipo de análisis, debemos tener un mínimo de 30 pulsos. Más tarde compararemos la muestra dubitada con cada una de las muestras indubitadas.

Los datos recogidos son los siguientes:

Muestra dubitada	
	ANÁLISIS
Mean pitch	149,543667
Stan dev	3,41033333
Number of pulses	34

Muestra indubitada 1	
	ANÁLISIS
Mean pitch	239,171833
Stan dev	4,508
Number of pulses	59

Muestra indubitada 2	
	ANÁLISIS
Mean pitch	137,243
Stan dev	2,39183333
Number of pulses	38

Muestra indubitada 3	
	ANÁLISIS
Mean pitch	229,4185
Stan dev	2,752
Number of pulses	77

Muestra indubitada 4	
	ANÁLISIS
Mean pitch	146,089667
Stan dev	1,61633333
Number of pulses	58

Resultados:

Muestra dubitada – Muestra indubitada 1

P-Value: 0,0000

Muestra dubitada – Muestra indubitada 2

P-Value: 0,0000

Muestra dubitada – Muestra indubitada 3

P-Value: 0,0000

Muestra dubitada – Muestra indubitada 4

P-Value: 0,0000

Según los resultados que nos proporciona el P-Value de las comparaciones de la vocal [a], todas las muestras, al estar por debajo del 0.05, son diferentes de la muestra dubitada.

Como los resultados del análisis anterior no nos han proporcionado los resultados esperados, vamos a realizar otra prueba. Vamos a tomar una palabra que se repita en todas las muestras y vamos a realizar el mismo análisis de Statdisk que hemos aplicado a las vocales anteriormente, para luego analizar su P-Value.

Los datos recogidos son los siguientes:

- Palabra “**Lingüística**”:

"Lingüística" - Muestra dubitada	
Mean pitch	186.697
Stan. Dev.	27.689
Number of pulses	62

"Lingüística" - Muestra indubitada 1	
Mean pitch	218.394
Stan. Dev.	57.597
Number of pulses	113

"Lingüística" - Muestra indubitada 2	
Mean pitch	124.028
Stan. Dev.	13.377
number of pulses	32

"Lingüística" - Muestra indubitada 3	
Mean pitch	262.390
Stan. Dev.	12.329
number of pulses	144

"Lingüística" - Muestra indubitada 4	
Mean pitch	165.325
Stan. Dev.	3.697
number of pulses	92

Resultados:

Muestra dubitada – Muestra indubitada 1

P-Value: 0,0000

Muestra dubitada – Muestra indubitada 2

P-Value: 0,0000

Muestra dubitada – Muestra indubitada 3

P-Value: 0,0000

Muestra dubitada – Muestra indubitada 4

P-Value: 0,0000

En este caso, tampoco nos proporciona los resultados esperados, ya que, según el P-Value de las comparaciones de la palabra “lingüística”, todas las muestras, al estar por debajo del 0.05, son diferente de la muestra dubitada.

A partir de los datos que nos proporcionan las tablas anteriores, vamos a realizar un análisis otro nuevo análisis, donde observemos la distancia euclídea. Esta es la distancia que se observa entre dos puntos. La fórmula que se aplica en este análisis es la siguiente:

$$d(p, q) = \sqrt{\sum_i^n (p_i - q_i)^2}$$

Donde p recogería los datos de la muestra dubitada y q los de la indubitada. Debemos restar al valor de la muestra dubitada el de la indubitada y, si nos da un número negativo, debemos normalizarlo, es decir, ponerlos en números positivos.

En este caso, vamos a realizar este análisis del valor medio del pitch, que es el valor más fiable, ya que son los valores de frecuencia absoluta que caracterizan la voz y nos ayudará a ver cuál es la muestra indubitada que más se acerca a la dubitada. Estas distancias las representaremos en una serie de tablas comparativas:

- Vocal [a]

Muestra dubitada - Muestra indubitada 1		
Vocal [a]	M-Dub	M-Ind1
Valor medio Pitch [a]	128,69	282,18
Distancia	-153,49	
Valor medio Pitch [á]	142,72	202,71
Distancia	-59,99	

Valor absoluto [a]	153,49
Valor absoluto [á]	59,99
Suma	213,48

Muestra dubitada - Muestra indubitada 2		
Vocal [a]	M-Dub	M-Ind2
Valor medio Pitch [a]	128,69	149,68
Distancia	-20,99	
Valor medio Pitch [á]	142,72	147,39
Distancia	-4,67	

Valor absoluto [a]	20,99
Valor absoluto [á]	4,67
Suma	25,66

Muestra dubitada - Muestra indubitada 3		
Vocal [a]	M-Dub	M-Ind3
Valor medio Pitch [a]	128,69	250,75
Distancia	-122,06	
Valor medio Pitch [á]	142,72	206,53
Distancia	-63,81	

Valor absoluto [a]	122,06
Valor absoluto [á]	63,81
Suma	185,87

Muestra dubitada - Muestra indubitada 4		
Vocal [a]	M-Dub	M-Ind4
Valor medio Pitch [a]	128,69	162,65
Distancia	-33,96	
Valor medio Pitch [á]	142,72	141,16
Distancia	1,56	

Valor absoluto [a]	33,96
Valor absoluto [á]	1,56
Suma	35,52

Como podemos comprobar en estas tablas, de la vocal [a], la distancia calculada del pitch que más se acerca a la muestra dubitada es la de la muestra indubitada 2.

- Vocal [o]

Muestra dubitada - Muestra indubitada 1		
Vocal [o]	M-Dub	M-Ind1
Valor medio Pitch [o]	175,63	227,24
Distancia	-51,61	
Valor medio Pitch [o]	145,08	214,5
Distancia	-69,42	

Valor absoluto [o]	51,61
Valor absoluto [o]	69,42
Suma	121,03

Muestra dubitada - Muestra indubitada 2		
Vocal [o]	M-Dub	M-Ind2
Valor medio Pitch [o]	175,63	121,67
Distancia	53,96	
Valor medio Pitch [o]	145,08	126,4
Distancia	18,68	

Valor absoluto [o]	53,96
Valor absoluto [o]	18,68
Suma	72,64

Muestra dubitada - Muestra indubitada 3		
Vocal [o]	M-Dub	M-Ind3
Valor medio Pitch [o]	175,63	246,96
Distancia	-71,33	
Valor medio Pitch [o]	145,08	230,49
Distancia	-85,41	

Valor absoluto [o]	71,33
Valor absoluto [o]	85,41
Suma	156,74

Muestra dubitada - Muestra indubitada 4		
Vocal [o]	M-Dub	M-Ind4
Valor medio Pitch [o]	175,63	143,7
Distancia	31,93	
Valor medio Pitch [o]	145,08	141,44
Distancia	3,64	

Valor absoluto [o]	31,93
Valor absoluto [o]	3,64
Suma	35,57

En el caso de la vocal [o], la distancia calculada del pitch que más se acerca a la muestra dubitada es la de la muestra indubitada 4.

- Vocal [u]

Muestra dubitada - Muestra indubitada 1		
Vocal [u]	M-Dub	M-Ind1
Valor medio Pitch [u]	153,86	241,83
Distancia	-87,97	
Valor medio Pitch [u] antes nasal	152,89	252,27
Distancia	-99,38	

Valor absoluto [u]	87,97
Valor absoluto [u] antes de nasal	99,38
Suma	187,35

Muestra dubitada - Muestra indubitada 2		
Vocal [u]	M-Dub	M-Ind2
Valor medio Pitch [u]	153,86	143
Distancia	10,86	
Valor medio Pitch [ü]	152,89	143,04
Distancia	9,85	

Valor absoluto [u]	10,86
Valor absoluto [ü]	9,85
Suma	20,71

Muestra dubitada - Muestra indubitada 3		
Vocal [u]	M-Dub	M-Ind3
Valor medio Pitch [u]	153,86	216,74
Distancia	-62,88	
Valor medio Pitch [u] antes de nasal	152,89	198,09
Distancia	-45,2	

Valor absoluto [u]	62,88
Valor absoluto [u] antes de nasal	45,2
Suma	108,08

Muestra dubitada - Muestra indubitada 4		
Vocal [u]	M-Dub	M-Ind4
Valor medio Pitch [u]	153,86	142,11
Distancia	11,75	
Valor medio Pitch [u] antes nasal	152,89	140,38
Distancia	12,51	

Valor absoluto [u]	11,75
Valor absoluto [u] antes nasal	12,51
Suma	24,26

Como en el caso de la [o], en la [u] la distancia calculada del pitch que más se acerca a la muestra dubitada es la de la muestra indubitada 2.

7.2.3. Análisis Lingüístico

Signature

Vamos a realizar comparaciones de números de letras y de palabras. El número de oraciones o el de signos de puntuación no podemos observarlo ya que el texto ha sido sacado de muestras orales. Por lo tanto, los signos de puntuación han sido colocados por los autores de este proyecto, como ellos han creído conveniente. No son las puntuaciones que a lo mejor pondrían los autores reales de estas muestras.

Después de recoger los resultados, debemos investigar la probabilidad de que los textos sean iguales o no. Si en alguno de los casos, la diferencia entre los textos es menor que 0.05 es que son diferentes, por el contrario, si se acerca más a 1, es que son totalmente iguales.

Muestra Dubitada – Muestra Indubitada 1

Muestra dubitada - Muestra indubitada 1			
nº de letras	Total	nº de palabras	Total
Muestra dubitada	599		103
Muestra indubitada 1	1011		189
M.Dub-M.Ind1	1610		292

- Chi-square significance test:

	Words	Letters
Difference	0.05	0.02
Chi-Square 2% value		31
1% value		33.41
Chi-Square 5% value	9.49	
2% value	11.67	

Muestra Dubitada – Muestra Indubitada 2

Muestra dubitada - Muestra indubitada 2			
nº de letras	Total	nº de palabras	Total
Muestra dubitada	599		103
Muestra indubitada 1	527		95
M.Dub-M.Ind1	1126		198

- Chi- square significance test:

	Words	Letters
Difference	0.001	0.2
Chi-Square 20% value		20.46
10% value		23.54
Chi-Square 0.1%	20.52	

Muestra Dubitada – Muestra Indubitada 3

Muestra dubitada - Muestra indubitada 3			
nº de letras	Total	nº de palabras	Total
Muestra dubitada	599		103
Muestra indubitada 1	784		149
M.Dub-M.Ind1	1383		252

- Chi- square significance test:

	Words	Letters
Difference	0.01	0.05
Chi-Square 5% value		25
2% value		28.26
Chi-Square 1% value	13.28	
0.1%	18.46	

Muestra Dubitada – Muestra Indubitada 4

Muestra dubitada - Muestra indubitada 1			
nº de letras	Total	nº de palabras	Total
Muestra dubitada	599		103
Muestra indubitada 1	761		150
M.Dub-M.Ind1	1360		253

- Chi- square significance test

	Words	Letters
Difference	0.2	0.02
Chi-Square 2% value		31
1% value		33.41
Chi-Square 20% value	5.99	
10% value	7.78	

Comparación de resultados:

M. Dubitada – M. Indubitada 1	Palabras	Letras
Diferencia	0.05	0.02

M. Dubitada – M. Indubitada 2	Palabras	Letras
Diferencia	0.001	0.2

M. Dubitada – M. Indubitada 3	Palabras	Letras
Diferencia	0.01	0.05

M. Dubitada – M. Indubitada 4	Palabras	Letras
Diferencia	0.2	0.02

Una vez realizada la comparación, podemos decir que en cuanto a las palabras, las comparaciones de la muestra dubitada con las muestras indubitadas 1, 2 y 3 se consideran diferentes, ya que su diferencia es menor o igual que 0.05. En cambio, la muestra indubitada

4 tiene una diferencia de 0.2, por lo que podríamos decir que tiene más posibilidades de compartir autoría, porque se acerca más a 1. Por otra parte, en cuanto a las letras, las comparaciones de la muestra dubitada con las muestras indubitadas 1, 3 y 4, se consideran diferentes, ya que su diferencia es menor o igual que 0.05, mientras que la muestra indubitada 2 se consideraría más parecida a la dubitada, porque su diferencia es de 0.2, es decir, más cercana a 1.

CopyCatch

En el análisis que vamos a realizar con CopyCatch, vamos observar los porcentajes de frecuencia de vocabulario común de todas las muestras. Se deberán aproximar más las muestras que tengan mayor probabilidad de ser de la misma persona, pero, como todo análisis, tiene un margen de error, por lo que podría fallar.

	Vocabulario común
M. Dubitada – M. Indubitada 1	20%
M. Dubitada – M. Indubitada 2	11%
M. Dubitada – M. Indubitada 3	12%
M. Dubitada – M. Indubitada 4	17%

Tras este análisis, la muestra indubitada que más vocabulario tiene en común con la muestra dubitada, es la número 1.

8. Dictamen

Después de haber realizado el análisis articulatorio, el acústico y el lingüístico, podemos observar, tras la comparación de los datos, las siguientes características:

1. Según los datos que nos proporciona el análisis articulatorio, tanto en el análisis de rasgos como en el de vulgarismos cometidos, la muestra que más se parece a la dubitada es la número 2, ya que es la que comparte más rasgos. Esto hace que sea la muestra con mayor grado de probabilidad de similitud, es decir, existen una alta probabilidad de que el autor de la muestra indubitada 2 coincida con el de la muestra dubitada.

Tablas comparativas y niveles de similitud:

Comparaciones	Número de veces que coinciden sus rasgos articulatorios
Muestra dubitada – Muestra indubitada 1	5
Muestra dubitada – Muestra indubitada 2	9
Muestra dubitada – Muestra indubitada 3	5
Muestra dubitada – Muestra indubitada 4	3

Niveles de similitud según los rasgos articulatorios				
	M.IND 1	M.IND 2	M.IND 3	M.IND 4
M.DUB	Inconclusión	Medio – alto	Inconclusión	Medio - bajo

Comparaciones	Número de veces que coinciden sus rasgos idiolectales o de registro
Muestra dubitada – Muestra indubitada 1	6
Muestra dubitada – Muestra indubitada 2	7
Muestra dubitada – Muestra indubitada 3	5
Muestra dubitada – Muestra indubitada 4	4

Niveles de similitud según los rasgos idiolectales o de registro				
	M.IND 1	M.IND 2	M.IND 3	M.IND 4
M.DUB	Inconclusivo	Medio – alto	Inconclusivo	Medio - bajo

2. Según los datos de los análisis acústicos, hemos podido observar que, analizando la medida de los formantes 1 y 2 de las vocales [a] en posición final dentro de la palabra “lingüística”, [á], [o] dentro del artículo “los”, [o] en posición final y dos [u] antes de [n], la muestra indubitada que más se asemeja a la muestra dubitada es la número 2.

En cuanto al análisis realizado mediante los test de hipótesis (Statdisk), en todos los casos el P-Value nos sale 0. Por lo tanto, a través de este test, podemos decir que todas las muestras son diferentes. No nos proporciona un resultado por el cual podamos sospechar que alguna de las muestras indubitadas sea del mismo autor que la muestra dubitada.

Para seguir intentando comprobar cuál de las muestras se asemeja más a la dubitada, hemos realizado, a partir de los rasgos acústicos, una operación para determinar la distancia euclídea que existe entre las vocales dependiendo de las muestras. Los resultados que nos proporciona nos informan de que la muestra dubitada está más cerca de la muestra indubitada 2 en las vocales [a] y [u] y de la muestra indubitada 4 en la vocal [o].

Tablas comparativas y niveles de similitud:

[a]	Distancia euclídea
M.DUB – M.IND 1	213,48
M.DUB – M.IND 2	25,66
M.DUB – M.IND 3	185,87
M.DUB – M.IND 4	35,52

Niveles de similitud según la distancia euclídea [a]

	M.IND 1	M.IND 2	M.IND 3	M.IND 4
M.DUB	Eliminación	Alta	Baja	Medio - alto

[o]	Distancia euclídea
M.DUB – M.IND 1	121,03
M.DUB – M.IND 2	72,64
M.DUB – M.IND 3	156,74
M.DUB – M.IND 4	35,57

Niveles de similitud según la distancia euclídea [o]

	M.IND 1	M.IND 2	M.IND 3	M.IND 4
M.DUB	Medio – bajo	Inconclusivo	Baja	Medio - alto

[u]	Distancia euclídea
M.DUB – M.IND 1	187,35
M.DUB – M.IND 2	20,71
M.DUB – M.IND 3	108,08
M.DUB – M.IND 4	24,26

Niveles de similitud según la distancia euclídea [u]

	M.IND 1	M.IND 2	M.IND 3	M.IND 4
M.DUB	Baja	Alta	Medio – bajo	Alta

3. Analizando los datos del análisis lingüístico, según el análisis del chi – square test, podemos decir que, en cuanto a palabras, la muestra indubitada más parecida a la muestra dubitada sería la número 4, mientras que, si nos fijamos en las letras, la muestra indubitada que más se parece es la número 2.

Tablas comparativas y niveles de similitud:

Comparaciones	Palabras	Letras
Muestra dubitada – Muestra indubitada 1	0.05	0.02
Muestra dubitada – Muestra indubitada 2	0.001	0.2
Muestra dubitada – Muestra indubitada 3	0.01	0.05
Muestra dubitada – Muestra indubitada 4	0.2	0.02

Niveles de similitud según el número de palabras	M.IND 1	M.IND 2	M.IND 3	M.IND 4
M.DUB	Inconclusivo	Eliminación	Eliminación	Medio - alto

Niveles de similitud según el número de letras	M.IND 1	M.IND 2	M.IND 3	M.IND 4
M.DUB	Eliminación	Medio – alto	Inconclusivo	Eliminación

Si miramos los resultados sobre el vocabulario que nos proporcionó CopyCatch, la muestra indubitada que más vocabulario tiene en común con la muestra dubitada, es la número 1.

Tablas comparativas y niveles de similitud:

Comparaciones	Porcentaje de vocabulario común
Muestra dubitada – Muestra indubitada 1	20%
Muestra dubitada – Muestra indubitada 2	11%
Muestra dubitada – Muestra indubitada 3	12%
Muestra dubitada – Muestra indubitada 4	17%

Niveles de similitud según los porcentajes de vocabulario común				
	M.IND 1	M.IND 2	M.IND 3	M.IND 4
M.DUB	Inconclusivo	Medio – bajo	Medio – bajo	Medio - bajo

Una vez observados todos los datos necesarios, y habiendo realizado todos los análisis pertinentes, debemos concluir que ninguna de las muestras indubitadas tiene un grado de similitud lo suficientemente alto como para poder afirmar que el autor de la muestra dubitada coincide con el autor de alguna muestra indubitada.

Tras recopilar y analizar todos los resultados obtenidos, podemos observar una mayor similitud de la muestra indubitada 2 con la muestra dubitada, ya que existe una mayor coincidencia en todos los análisis realizados. Pero ninguno de ellos nos garantiza que un mismo individuo sea el locutor de la muestra dubitada y de alguna de las indubitadas.

Para llegar a una buena identificación deberíamos realizar una serie de análisis más exhaustivos e investigar más a fondo en todas las muestras. Solo analizando acústicamente las vocales fuertes, como hemos hecho en este trabajo, no es suficiente para saber con certeza la autoría de dichas muestras. Debemos observar por ejemplo, la medida del resto de vocales, la longitud del sonido [s] final o la medida de los formantes de las consonantes laterales. Todo esto nos proporcionaría muchísimos más datos para poder comparar las muestras y, seguramente, nos daría una probabilidad de similitud más alta. Pero para llevar a cabo este tipo de investigación debemos contar con un trabajo mucho más amplio. En este proyecto lo que hemos analizado han sido partes básicas de las muestras para realizar el simulacro de peritaje del que ya habíamos hablado anteriormente. A la hora de realizar un análisis real, como es el caso de los realizados por un perito, se deberían hacer todos los análisis posibles de todos los rasgos característicos para hallar la muestra que comparta autoría con la muestra dubitada.

9. Bibliografía

ASOCIACIÓN INDEPENDIENTE DE PERITOS JUDICIALES (2015) <http://judiciales.es/>

Black's Law Dictionary 6th ed. (West Publishing, 1990)

CASALES, F. (2005). Las variedades diafásicas en el discurso glotodidáctico. *Revista de estudios literarios*. Universidad Complutense de Madrid, Espéculo.

CERNUDA DEL RÍO, A. (2001). *Informática- Microsoft Excel* (I). Versión 1.0.

COULTHARD, M. and JOHNSON, A.. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*. London and New York: Routledge.

BOBADILLA, J. GÓMEZ, P, BERNAL, J. (1999). *Posición y evolución de los formantes del habla. Estado del arte*. En *Estudios de Fonética Experimental X* (13 - 40). Barcelona: Laboratorio de Fonética.

DELGADO ROMERO, C. (2001): “La identificación de locutores en el ámbito forense”, Universidad Complutense de Madrid (Tesis doctoral).

DELGADO ROMERO, C. (2012). *La pericia de identificación de habla: El papel fundamental del experto*. En *Lingüística Forense: La Lingüística en el ámbito legal y policial*.(101-114). Madrid: Euphonía Ediciones.

FERNÁNDEZ PLANAS, A.M. (2005). *Así se habla. Nociones fundamentales de fonética general y española*. Barcelona: Horsori.

FRÍAS CONDE, X. (2001). *Introducción a la Fonética y Fonología del Español*. *Ianua. Revista Philologica Romanica*, 04, 3-23.

GONZÁLEZ SIGÜENZA, B. (2008). *BATVOX: Sistema automático de reconocimiento de locutor*, 301 – 316. Agtínio.

HALLIDAY, Michael Alexander Kirkwood, Angus MCINTOSH y Peter STREVENS. (1964). *The Linguistic Sciences and Language Teaching*. Londres: Longmans.

International Phonetic Alphabet. Promoting The Study Of Phonetics.
<http://www.internationalphoneticalphabet.org/ipa-sounds/ipa-chart-with-sounds/>

JIMÉNEZ BERNAL, M; REIGOSA RIVEIROS, M; GARAYZÁBAL HEINZE, E. (2010): *Lingüista Forense: Licencia para investigar la lengua*. En Garayzábal, Jiménez y Reigosa (Eds.) Panorama actual de la Lingüística Forense en el ámbito legal y policial. Teoría y práctica. Euphonia Ediciones. Madrid.

JIMÉNEZ BERNAL, M., REIGOSA RIVEIROS, M., GARAYZÁBAL HEINZE, E. (2012). *La Lingüística Forense: La licencia para investigar la lengua*. En Lingüística Forense. La Lingüística en el ámbito legal y policial. (19-39). Madrid: Euphonía Ediciones.

JIMÉNEZ FERNÁNDEZ, R. (1999). El andaluz. Madrid: Arco Libros.

Legislación consolidada. Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal.
<http://www.boe.es/buscar/act.php?id=BOE-A-1995-25444>

LUCENA MOLINA, J.J. (2005). *La acústica forense*. 2015, de Instituto Universitario sobre seguridad interior. Universidad Nacional de Educación a distancia. Sitio web:
http://portal.uclm.es/descargas/idp_docs/doctrinas/referencia_autores_guardia_civil_i_u_i_s_i.pdf

MARTÍNEZ CELDRÁN, E. (1991). *Fonética experimental: Teoría y práctica*. Madrid: Síntesis.

MAYORAL ASENSIO, Roberto (1998). *La traducción de la variación lingüística*. Tesis doctoral de la Universidad de Granada.

MODÉJAR, J. (1991). V. El andaluz: visión de conjunto. En Dialectología Andaluza(121 - 145). Málaga: Analecta Malacitana.

NARBONA, A., CANO, R., MORILLO, R. (1998). El español hablado en Andalucía. Barcelona: Ariel S.A.

QUILIS, A. (1999). *Tratado de Fonética y Fonología Españolas*. Leganés (Madrid): Gredos.

QUILIS, A. (19982): *Principios de fonología y fonética españolas*. Madrid: Arcolibros.

ROMÁN MONTES DE OCA, D. (2004). *Manual de Praat*. 2015, de Laboratorio de Fonética. Facultad de Letras. Pontificia Universidad Católica de Chile. Sitio web: http://domingo-roman.net/manual_analisis_acustico.html.

ROMÁN MONTES DE OCA, D. (2008). *Recursos para el análisis acústico de la melodía del habla en praat*, 343 – 157. Pontificia Universidad Católica de Chile.

SÁNCHEZ GÓMEZ, M.a C. y REVUELTA DOMÍNGUEZ, F.I. (2005). *El proceso de transcripción en el marco de la metodología de investigación cualitativa actual*. *Enseñanza*, 23, 367-386.

SERVICIO DE TÉCNICA POLICIAL. SERVICIO DE ACÚSTICA FORENSE. Informe pericial de análisis de voz. Dirección general de policía, Comisaría general de policía científica.

SÁNCHEZ RÍOS, J.A. (2002). *Validación de parámetros acústicos para la voz expresiva*. Universitat Autònoma de Barcelona. Facultat de Ciències de la comunicació. Departament de comunicació audiovisual i de publicitat. (Tesis doctoral).

ROSE, P. (2002). *Forensic Speaker Identification*. London and New York: Taylor & Francis.

SHUY, R. (2000). *Forensic Linguistics*. En Aronoff, Mark and Janie Rees – Millers, eds. *The handbook of Linguistics*. Oxford: Blackwell.

TURELL, M. Teresa (ed) (2005). *Lingüística Forense, Lengua y Derecho: Conceptos, Métodos y Aplicaciones*. Barcelona: IULA.

TURELL, M. T. (2006). "¡Temblad, plagiarios, llega el CopyCatch!". *La Vanguardia*, 1.

10. Anexos

Adjuntamos las muestras una vez realizada la transcripción. Solo se ha realizado la transcripción de aquellas partes que han sido útiles para llevar a cabo nuestro estudio.

- **Muestra dubitada**

A: ¿Cuál es tu carrera?

B: [e:]h... Lingüí[S]tica y Lengua[S] Aplicada[S]

A: ¿Y por qué la elegiste?

B: (Coge aire fuerte) pu[e:] [S] porqu[e:]... ay, no se (risas) no, si, porque me interesan mucho lo[S] idioma[S] y... bueno, por otra parte ne[h]ecitaba... era con[S]ciente que ne[h]ecitaba u[n:] título universitario y decidí que... me parecía mucho más interesante hacer... [e:]h.. hacer Lingüí[S]tica. Me pareci[ó:] interesante hacer e[S]ta carrera par[a:] complementar un... un poco lo[S] estudio[S] de inglés que ya tengo, con lo[S] que me dedico [a:] dar clase a niño[S] pequeño[S]... Primero pensé hacer... hacer otra filolo[G]ía para especializarme e[n:]... francé[S] por si ne[h]esitara otro idioma má[S] en un futuro pero... bueno, me decidí por... por e[S]ta.

- **Muestra indubitada 1**

A: ¿Cuál es tu carrera?

B: Pu[e:]s...estoy cursando lingüí [h]stica

A: Y ¿Por qué te decidiste por ella?

B: (risas)... si te contara el porqué, bueno, pue[h] me remonto, en el dos mil nueve acabé el ba[ʃ]illerato y.. y... y oposité a l[a:]... a la guardia civil, en lo[h] do[h] año[h] que e[h]tuve opositando no conseguí nin... plaza pero si e[h]tab[a:] clasificada. Total, que luego decidí meterme en dere[ʃ]o..... e[h]tuve ha[h]ta el mes d[e:] noviembre en li[h]ta de e[h]pera (aguanta la risa), me pareció super guay... con to' la ilusión del mundo de que iba a entrá en dere[ʃ]o, pu[e:]s no entré... e[l:].. me ofrecieron (risas) una serie d[e:], una serie de.... la[h] carrera[h] que e[h]taban, que había plazas... sue[r]tas y miré que...quedaba lingüí[h]tica y digo pos'... ¿de qué va e[S]to? me metí en la pá[G]ina de la UCA y vi lo primero Lingüí[h]tica Forense y yo creí que... lingüí[h]tica forense era lo primero que iba a dar (risas), (coge aire fuerte) y me metí en lingüística, no era m[á:].... nada, al principio me creí que era una [ʃ]orrada de.. de carrera realmente, pero... má[h] o meno[h] luego ya fui conociendo la[h] asignatura[h] y digo, po' nada, me voy a quedar en ella, sino me hubiera cambio'.

- **Muestra indubitada 2**

A: ¿Cuál es tu carrera?

B: eh, Lingüística y lengua aplicada

A: Y ¿Por qué te decidiste por ella?

B: (coge aire fuerte) porqu[e:]... bueno, mi primera opción realmente fue informática per[o:]... puse e[h]ta como segunda porque me gu[h]tan mu[cho] lo[h] idioma[h] y me gu[h]taría dedicarme[e:].. a tradu[C]ción o algo parecid[o:]... relacionado con lo[h] idioma[h]. Por otra part[e:] también se me han dado bien la[h] letra[h] de[h] de siempre, por es[o:] la puse com[o:] segunda opción, aunque[e:]... bueno, prefiero la informátic[a:] [e:]... per[o:] tenía que tener un[a:]... una opción por s[i:] no entraba en la primera... Al principi[o:] no me convencía demasiao', per[o:] poco a poco me fu[e:] gu[h]tando.

- **Muestra indubitada 3**

A: ¿Cuál es tu carrera?

B: Mi carrera [e: h] Lingüística y lengua [h] Aplicada[h]

A: Y ¿Por qué te decidiste por ella?

B: Pu[e:][h] la ele[h]í... en principio no lo tenía muy claro, n[o:] sabía que e[h] lo que iba a estudiá per[o:] en el in[h]tituto se me daban bastante mejó la[h] letra[h] que la[h] ciencia[h].. y.. como no quería desplazarme de Cádi p[o:]' elegí e[h]ta carrera, que al principio no m[e:] result[ó:] lo que yo quería per[o:] de[h]pué[h] como tambié[n:] tenemo[h] alguna[h] asi[h]natura[h] que están más relacioná[h] con lo[h] idioma[h] y... se me da un poco mejó pu[é:][h] la verdá e[h] que me h[a:]... me ha ido gu[h]tando un poco má[h].. y luego, al principio po'[h] como to' la[h] carrera[h] la[h] primera[h] asi[h]natura[h] sinta[S]i[h], fonética, morfolo[h]ía me parecieron un poco má[h] rollo, pero luego cuando va[h] llegand[o:] a curso'superior[e:][h]'n tercero en cuarto p[o:]' ya cuando llega Lingüística Clínic[a:], Lingüística Forens[e:]... mol[a:] má[h].

- **Muestra indubitada 4**

A: ¿Cuál es tu carrera?

B: Pu[e:] yo e[h]toy e[h]tudiando Lingüí[h]tica y Lengua[h] Aplicada[h].

A: Y ¿Por qué te decidiste por ella?

B: Pu[e:] [h] en verd[á:] yo co[h]í e[h]ta carrera d[e:] rebote porqu[e:] me metí en Filolo[h]ía árabe pero la vi demasiao' complicá' enton[s]e p[o:]' a[r]go que fuera en e[^h]pañó' y un poquito m[á:] [h] sen[s]illo, que a[r] finá' no lo ha sido, ha sio' Lingüí[h]tica y Lengua[h] Aplicada, y nada en verd[á:]' e[h]tá bien, sobre to' l[a:] [h] ú[r]tima[h] así [h]natura[h] que hemo[h] dao' ya lo que e[h] el... el... el e[h]]pe[s]ífico qu[e:] e[h] Lingüí[h]tica Clínica, Lingüí[h]ti Fo... Lingüí[h]tica Forense y eso qu[e:] e[h] lo que má[h] me e[h]tá llamando la aten[s]ión, porque lo que e[h] la morfolo[h]ía... la sinta[s]i'y to es[o:] la verdá e[h] que no se me da ni bien (coge aire con fuerza) y... nada, eso, ahora bu[h]cando un má[h]te' p[a:]... pa' completá' lo[h] e[h]tudio[h] y... a vé' si puede sé' de Lingüí[h]tica Clínica... adió'.

Leyenda de la transcripción:

[vocal o consonante final :] → Alargamiento de vocal o consonante final.

[h] → Indica aspiraciones.

[ħ] → Glotalización de [x].

[S] → [s] muy marcada.

‘ → Marca las palabras incompletas y las sinalefas.

[G] → [g] muy marcada.

Análisis signature

Muestra Dubitada – Muestra Indubitada 1

- Chi-square significance test:

Words:

Some data columns, with expected values less than 5, have been ignored in this calculation

Columns:	5	Degrees of freedom:	4	Chi-square:	6,868
----------	---	---------------------	---	-------------	-------

The aim of this test is to assess the significance of the measured differences between the 'sample' and the 'reference' texts.

In this case the difference is significant only at the 20% (0.2) level.
Chi-Square 20% value = 5,99; 10% value = 7,78

Word lengths Sentence lengths Paragraph lengths Letters Punctuation

Letters:

Some data columns, with expected values less than 5, have been ignored in this calculation

Columns:	18	Degrees of freedom:	17	Chi-square:	31,687
----------	----	---------------------	----	-------------	--------

The aim of this test is to assess the significance of the measured differences between the 'sample' and the 'reference' texts.

In this case the difference is significant at the 2% (0.02) level.
Chi-Square 2% value = 31; 1% value = 33,41

Word lengths Sentence lengths Paragraph lengths Letters Punctuation

Muestra Dubitada – Muestra Indubitada 2

- Chi-square significance test:

Words:

Some data columns, with expected values less than 5, have been ignored in this calculation

Columns:	5	Degrees of freedom:	4	Chi-square:	6,868
----------	---	---------------------	---	-------------	-------

The aim of this test is to assess the significance of the measured differences between the 'sample' and the 'reference' texts.

In this case the difference is significant only at the 20% (0.2) level.
Chi-Square 20% value = 5,99; 10% value = 7,78

Word lengths	Sentence lengths	Paragraph lengths	Letters	Punctuation
--------------	------------------	-------------------	---------	-------------

Letters:

Some data columns, with expected values less than 5, have been ignored in this calculation

Columns:	18	Degrees of freedom:	17	Chi-square:	31,687
----------	----	---------------------	----	-------------	--------

The aim of this test is to assess the significance of the measured differences between the 'sample' and the 'reference' texts.

In this case the difference is significant at the 2% (0.02) level.
Chi-Square 2% value = 31; 1% value = 33,41

Word lengths	Sentence lengths	Paragraph lengths	Letters	Punctuation
--------------	------------------	-------------------	---------	-------------

Muestra Dubitada – Muestra Indubitada 3

- Chi-square significance test:

Words:

Some data columns, with expected values less than 5, have been ignored in this calculation

Columns:	5	Degrees of freedom:	4	Chi-square:	6,868
----------	---	---------------------	---	-------------	-------

The aim of this test is to assess the significance of the measured differences between the 'sample' and the 'reference' texts.

In this case the difference is significant only at the 20% (0.2) level.
Chi-Square 20% value = 5,99; 10% value = 7,78

Word lengths	Sentence lengths	Paragraph lengths	Letters	Punctuation
--------------	------------------	-------------------	---------	-------------

Letters:

Some data columns, with expected values less than 5, have been ignored in this calculation

Columns:	18	Degrees of freedom:	17	Chi-square:	31,687
----------	----	---------------------	----	-------------	--------

The aim of this test is to assess the significance of the measured differences between the 'sample' and the 'reference' texts.

In this case the difference is significant at the 2% (0.02) level.
Chi-Square 2% value = 31; 1% value = 33,41

Word lengths	Sentence lengths	Paragraph lengths	Letters	Punctuation
--------------	------------------	-------------------	---------	-------------

Muestra Dubitada – Muestra Indubitada 4

- Chi-square significance test:

Words:

Some data columns, with expected values less than 5, have been ignored in this calculation

Columns:	5	Degrees of freedom:	4	Chi-square:	6,868
----------	---	---------------------	---	-------------	-------

The aim of this test is to assess the significance of the measured differences between the 'sample' and the 'reference' texts.

In this case the difference is significant only at the 20% (0.2) level.
Chi-Square 20% value = 5,99; 10% value = 7,78

Word lengths	Sentence lengths	Paragraph lengths	Letters	Punctuation
--------------	------------------	-------------------	---------	-------------

Letters:

Some data columns, with expected values less than 5, have been ignored in this calculation

Columns:	18	Degrees of freedom:	17	Chi-square:	31,687
----------	----	---------------------	----	-------------	--------

The aim of this test is to assess the significance of the measured differences between the 'sample' and the 'reference' texts.

In this case the difference is significant at the 2% (0.02) level.
Chi-Square 2% value = 31; 1% value = 33,41

Word lengths	Sentence lengths	Paragraph lengths	Letters	Punctuation
--------------	------------------	-------------------	---------	-------------

Statdisk

Comparaciones de vocales:

Muestra dubitada – Muestra indubitada 1

Not eq. vars: No Pool (and df calculated with Formula 9-1)

Alternative Hypothesis:

μ_1 not equal μ_2

Test Statistic, t: -108,1785

Critical t: ± 1.988506

P-Value: 0,0000

Degrees of freedom: 84,2954

95% Confidence interval:

$-91.27552 < \mu_1 - \mu_2 < -87.98048$

Muestra dubitada – Muestra indubitada 2

Not eq. vars: No Pool (and df calculated with Formula 9-1)

Alternative Hypothesis:

μ_1 not equal μ_2

Test Statistic, t: 17,5277

Critical t: ± 2.001461

P-Value: 0,0000

Degrees of freedom: 58,3484

95% Confidence interval:

$10.89548 < \mu_1 - \mu_2 < 13.70452$

Muestra dubitada – Muestra indubitada 3

Not eq. vars: No Pool (and df calculated with Formula 9-1)

Alternative Hypothesis:

μ_1 not equal μ_2

Test Statistic, t: -120,3668

Critical t: ± 2.005909

P-Value: 0,0000

Degrees of freedom: 52,8137

95% Confidence interval:

$-81.20611 < \mu_1 - \mu_2 < -78.54389$

Muestra dubitada – Muestra indubitada 4

Not eq. vars: No Pool (and df calculated with Formula 9-1)

Alternative Hypothesis:
 μ_1 not equal μ_2

Test Statistic, t: 5,5520
Critical t: ± 2.018307
P-Value: 0,0000

Degrees of freedom: 41,8411

95% Confidence interval:
 $2.198379 < \mu_1 - \mu_2 < 4.709621$

Comparaciones de palabras:

Muestra dubitada – Muestra indubitada 1

Not eq. vars: No Pool (and df calculated with Formula 9-1)

Alternative Hypothesis:
 μ_1 not equal μ_2

Test Statistic, t: -4,9071
Critical t: ± 1.973962
P-Value: 0,0000

Degrees of freedom: 170,6369

95% Confidence interval:
 $-44.44755 < \mu_1 - \mu_2 < -18.94645$

Muestra dubitada – Muestra indubitada 2

Not eq. vars: No Pool (and df calculated with Formula 9-1)

Alternative Hypothesis:
 μ_1 not equal μ_2

Test Statistic, t: 15,2200
Critical t: ± 1.985546
P-Value: 0,0000

Degrees of freedom: 93,9111

95% Confidence interval:
 $54.49344 < \mu_1 - \mu_2 < 70.84456$

Muestra dubitada – Muestra indubitada 3

Not eq. vars: No Pool (and df calculated with Formula 9-1)

Alternative Hypothesis:

μ_1 not equal μ_2

Test Statistic, t: -20,6613

Critical t: ± 1.993635

P-Value: 0,0000

Degrees of freedom: 71,6361

95% Confidence interval:

$-82.99673 < \mu_1 - \mu_2 < -68.38927$

Muestra dubitada – Muestra indubitada 4

Not eq. vars: No Pool (and df calculated with Formula 9-1)

Alternative Hypothesis:

μ_1 not equal μ_2

Test Statistic, t: 6,0414

Critical t: ± 1.998671

P-Value: 0,0000

Degrees of freedom: 62,4685

95% Confidence interval:

$14.30157 < \mu_1 - \mu_2 < 28.44243$